

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 02:17:46 : Search time 1792.55 Seconds  
(without alignments)  
10063.136 Million cell updates/sec

Title: US-09-640-211a-2076

Perfect score: 862  
Sequence: 1 caaacgtctccgtctctc.....gcttaaaaaaaaaaaaaa 862

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_cm: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htgo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result Query #  
No. Score Match Length DB ID Description

1	240.6	27.9	1033	8	AB005888	AB005888 Arabidops
2	237	27.5	1022	8	PSMYB26	Y11105 P. sativum m
3	231.8	26.9	645	8	AF175987	AF175987 Arabidops
4	223	25.9	872	8	AB058642	AB058642 Lilium hy
5	192.8	22.4	1343	8	AY026332	AY026332 Oryza sat
6	176.4	20.5	1295	8	CP033917	U33917 Craterostig
7	167.8	19.5	1423	8	AF262733	AF262733 Arabidops
8	166.4	19.3	13342	8	AF003140	AF003140 Oryza sat
9	166.4	19.3	145491	8	AP002883	AP002883 Oryza sat
10	162.2	18.8	1140	8	AF334815	AF334815 Arabidops
11	155.4	18.0	931	8	AY008377	AY008377 Arabidops
12	151.4	17.6	1033	8	AF034133	AF034133 Gossypium
13	143.2	16.6	1513	8	AF427146	AF427146 Zea mays
14	143.2	16.6	1601	8	ZM057002	U57002 Zea mays P
15	143.2	16.6	1802	8	MZEPFR	M73028 Zea mays pr
16	138.2	16.0	1176	8	D88617	D88617 Oryza sativ
17	136.8	15.9	1202	8	OSMYB1202	Y11414 O. sativa m
18	136.8	15.9	1270	8	HVMYB3	X70878 H. vulgare m
19	136.2	15.8	2162	8	PPPSRNA	X67050 P. patens m
20	134.8	15.6	489	8	AB044084	AB044084 Trifolium
21	134.2	15.6	1024	8	AP062894	AP062894 Arabidops
22	133.8	15.5	1070	8	AT062743	U62743 Arabidops
23	133.8	15.5	126599	2	AP003816	AP003816 Oryza sat
24	133.8	15.5	131531	2	AP003813	AP003813 Oryza sat
25	133.6	15.5	2352	6	AR164405	AR164405 Sequence
26	133.6	15.5	2373	8	OSCAMYB	X98355 O. sativa m
27	133.4	15.5	1032	8	AF371974	AF371974 Arabidops
28	133.4	15.5	1069	8	D88618	D88618 Oryza sativ
29	133.4	15.5	1512	8	AY063939	AY063939 Arabidops
30	133	15.4	1050	8	AF175989	AF175989 Arabidops
31	132.8	15.4	1235	8	AY060588	AY060588 Arabidops
32	132.8	15.4	1284	8	HVMYB1	X70877 H. vulgare m
33	132.8	15.4	1334	8	AF062864	AF062864 Arabidops
34	132.6	15.4	1102	8	HVMYB2	X70876 H. vulgare m
35	131.6	15.3	2287	8	ASA133638	AF133638 Avena sat
36	131	15.2	1220	8	AF161711	AF161711 Pimpinell
37	130.6	15.2	819	8	AB029165	AB029165 Glycine m
38	130.4	15.1	1869	8	AF114162	AF114162 Lolium te
39	130.2	15.1	1259	8	AF249310	AF249310 Arabidops
40	130.2	15.1	1355	8	OSMYB1355	Y11415 O. sativa m
41	130	15.1	2220	6	AR164404	AR164404 Sequence
42	130	15.1	2263	8	HVRNAGAM1	X87690 H. vulgare m
43	129	15.0	795	8	AB029160	AB029160 Glycine m
44	128.6	14.9	1135	8	AF401220	AF401220 Fragaria
45	128	14.8	1606	8	AF034132	AF034132 Gossypium

## ALIGNMENTS

RESULT 1  
AB005888  
LOCUS AB005888 1033 bp mRNA linear PLN 05-FEB-1999  
DEFINITION Arabidopsis thaliana mRNA for ATMWB3, complete cds.  
ACCESSION AB005888  
VERSION AB005888.1 GI:2280527  
KEYWORDS ATMWB3.  
SOURCE Arabidopsis thaliana (strain: Columbia) cDNA to mRNA.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 1033)  
REFERENCE  
AUTHORS Noji, M.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1997) Masaaki Noji, Chiba University, Faculty of  
Pharmaceutical Sciences, Yayoi-cho 1-33, Inage-ku, Chiba  
263, Japan (E-mail: mnojie@ph.chiba-u.ac.jp, Tel: +81-43-290-2906,  
Fax: +81-43-290-2905)  
2 (sites)  
REFERENCE Noji, M., Urao, T., Shinozaki, K.Y. and Shinozaki, K.  
AUTHORS Molecular cloning of two cDNAs encoding novel myb homologs from  
TITLE Arabidopsis (Accession Nos. AB005888 and AB005889) (PCR98-111)

JOURNAL FEATURES	Plant Physiol. 117, 720 (1999)
SOURCE	Location/Qualifiers
gene	1. 1033
CDS	/organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" 115. .795 /gene="Atmyb3" 115. .795 /gene="Atmyb3" /codon_start=1 /product="ATMYB3" /protein_id="BA21618.1" /db_xref="gi:2280528" /transaltion="MEKRGCGSSCGSSCAEAYRKGPWTWEEDLLILINYIANHGDEV WNSLAKSAGLKRKTSQRLRLNLTLPDRGRGNTPEQLIMELHAKGNRWKILAK HAMPTDNEIKNFCRTRIQKTIKOSDYTTSSVGSHTSETINDOASTSSHNVCDD OAMETSYPTSYOHTNMEFNYSNAATATATVDYVPMKIVDDQGTENTWGMDDIMS SMHLNCG"
BASE COUNT	362 a 182 c 208 g 281 t
ORIGIN	

RESULT	2
LOCUS	PSMYR26
DEFINITION	P.sativum mRNA for Myb-like protein (Myb26).
ACCESSION	Y11105
VERSION	Y11105.1
KEYWORDS	GI:1841474
SOURCE	Myb-like protein: Myb26.
ORGANISM	pea.
	<i>Plasum sativum</i>
	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; <i>Plisum</i> .
REFERENCE	1 (bases 1 to 1022)
AUTHORS	Ulmairi, A. and Strommer, J.
TITLE	Myb26: a MYB-like protein of pea flowers with affinity for promoters of phenylpropanoid genes
JOURNAL	Plant J. 12 (6), 1273-1284 (1997)
MEDLINE	98112025
REFERENCE	2 (bases 1 to 1022)
AUTHORS	Strommer, J.N.

TITLE	JOURNAL	FEATURES	SOURCE	CDS	BASE COUNT	ORIGIN
Direct Submission	Submitted (05-FEB-1997)	Molecular Biology and Genetics	Location/Qualifiers		370 a	
<i>J. N. Strommer, University of Guelph,</i>			1..1032		170 c	321 t
<i>Ontario N1G 2W1, CANADA</i>			/organism="Pisum sativum"			
			/db_xref="taxon:3888"			
			/tissue_type="flower bud"			
			92..745			
			/codon_start=1			
			/product="Myb26"			
			/protein_id="CA71992.1"			
			/db_xref="GI:1841475"			
			/db_xref="SPRMBL:P93474"			
			/translation="MDKKPCNSSDPEVRKGPWTMEEDLILIN IAHNGEYVNSLAK			
			AAGCKRTGSCRLMLYLRPDVRGNTTPTEOLLIMELHSGKGNRSKSLAKHLRGST			
			DNEIKNMPRTKILKIQYNDPNQONQOKSLFINHHHHHHHPDSSQSYNLYPBM			
			ETYSPTSGCTLEPPPTPTPIINDHHHNSCCANDNNNNNTYMSMEDTWSQOLLNGD"			

RESULT	3				
LOCUS	AF175987				
DEFINITION	Arabidopsis thaliana putative transcription factor (MYB24)	645 bp	mRNA	linear	PLN 30-AUG-2001
ACCESSION	AF175987				
VERSION	AF175987.1				
KEYWORDS					
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eumariophyta; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 645)				
AUTHORS	Kranz,H.D., Denekamp,M., Greco,R., Jin,H., Leyva,A., Meissner,R.C.,				
	Petroni,K., Urzainqui,A., Beyer,M., Martin,C., Smeekens,S.,				
	Tonelli,C., Paz-Ares,J. and Weisshaar,B.				
TITLE	Towards functional characterisation of the members of the R2R3-MYB				
JOURNAL	gene family from Arabidopsis thaliana				
	The Plant Journal : for cell and molecular biology. 16 (2), 263-276				
	(1998)				
MEDLINE	99056848				

PUBMED 9839469  
 REFERENCE 2 (bases 1 to 645)  
 AUTHORS Stracke,R., Werber,M. and Weisshaar,B.  
 TITLE The R2R3-MYB gene family in *Arabidopsis thaliana*  
 JOURNAL Curr. Opin. Plant Biol. 4, 447-456 (2001)  
 REFERENCE 3 (bases 1 to 645)  
 AUTHORS Stracke,R. and Weisshaar,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-1999) Dept. Biochemie, Max-Planck-Institut fuer Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln D-50829, Germany

FEATURES  
 source  
 1..645  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /note="ecotype: Columbia"  
 1..645  
 /gene="MYB24"  
 1..645  
 /gene="MYB24"  
 /note="R2R3-MYB family member"  
 /codon\_start=1  
 /product="putative transcription factor"  
 /protein\_id="AAD53092.1"  
 /db\_xref="GI:5823307"  
 /translation="MEKRESSGSGSGDAEVRKGPMTMEDLLINTYIANHGGVYVNS  
 LAKSAGIKRRTGKSCRMLNLYRPVRRGNTPEQOLLIMELHDKWRKSLAKHL  
 GRINDIRKFRRTKIQKTIKSGETTTGSSOSEPTINHHATISHYANDTOETMDKSP  
 TTSYOHASHINQNLNGNVPSGSLMPLVSQSEQNTWSVDLPMNITNGN"

BASE COUNT 219 a 125 c 153 g 148 t  
 ORIGIN

Query Match 26.9%; Score 231.8; DB 8; Length 645;  
 Best Local Similarity 81.3%; Pred. No. 9.7e-55;

Matches 269; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 64 tcccaagatcgaagtgagaaagccgtgagacgatgaaagaagatcctccatc 123  
 DB 34 TCAGGAGATCCAGAGGTGAAGAAAGGCCCATGACATGAAAGATTTGATTCTC 93  
 QY 124 aactacaatagaacacgagcgaagcttggaactccctagcaaaagctgctgctca 183  
 DB 94 AATTATATCCCAATCATGTGTAAGGAGTGTGGAACCTCTCCCAAAATCTGCAGAC 153  
 QY 184 aaacgtaacccggaagctgctgctgctgctgctgctgctgctgctgctgctgct 243  
 DB 154 AAACGCAACCGGAAAGTTGCCGCTCCGCTGCTGAATACCTCCGACCTGATGCGCA 213  
 QY 244 agaggaacaactcaactacagagcagctcctgatactgaaactgacatgcaagtgagg 303  
 DB 214 CCGGCAAAATATTCACACCAAGAACAGCTACATCATGAACTTCAATAATGGGGA 273  
 QY 304 aacaggtgctcaaaattgcaaacgactcctccggaagacatgacaatgataaagaac 353  
 DB 274 AATAGTGTGTCACAAATATTCGCAAGCAATTTACAGAGAACGACATGAGATAAAGAA 333  
 QY 364 ttctggagagactagaatcccaaaagcaatca 394  
 DB 334 TTTTGGAGACTAAGATCCAGAAATACATCA 364

RESULT 4  
 AB058642 872 bp mRNA linear PLN 04-APR-2001  
 LOCUS AB058642  
 DEFINITION Lilium hybrid division I Lhmyb mRNA, complete cds.  
 ACCESSION AB058642  
 VERSION AB058642.1 GI:13537529  
 KEYWORDS  
 SOURCE Lilium hybrid division I (cultivar:Montreux) tepal cDNA to mRNA.  
 ORGANISM Lilium hybrid division I  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae;  
 Lilium; Lilium hybrid cultivars.

REFERENCE 1 (sites)  
 AUTHORS Nakatsuka,A., Izumi,Y. and Yamagishi,M.  
 TITLE Isolation and characterization of the genes related to anthocyanin biosynthesis in Asiatic hybrid lily  
 JOURNAL unpublished  
 REFERENCE 2 (bases 1 to 872)  
 AUTHORS Nakatsuka,A., Izumi,Y. and Yamagishi,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAR-2001) Akira Nakatsuka, Shimane University, Faculty of Life and Environmental Sciences, Nishiwatsu 1060, Matsue, Shimane 690-8504, Japan  
 (E-mail:nakira@life.shimane-u.ac.jp, Tel:01-852-32-6502)

FEATURES  
 source  
 1..872  
 /organism="Lilium hybrid division I"  
 /cultivar="Montreux"  
 /db\_xref="taxon:156532"  
 /tissue\_type="tepals"  
 80..634  
 /gene="Lhmyb"  
 80..634  
 /gene="Lhmyb"  
 /codon\_start=1  
 /protein\_id="BAB40790.1"  
 /db\_xref="GI:13537530"  
 /translation="MDRKVGTENDEVEVRKGPMTMEDLLINTYIANHGGVYVNS  
 AGLRKTSCLRMLNLYRPVRRGNTPEQOLLIMELHDKWRKSLAKHLGRD  
 NEIKMFRTRVQKSKNGSESESGDISLSDTSAMQEDVDVGSOTSTDGQTSNNL  
 DAFETPEPBNLLSVEELWMOQ"

BASE COUNT 269 a 177 c 213 g 213 t  
 ORIGIN

Query Match 25.9%; Score 223; DB 8; Length 872;  
 Best Local Similarity 69.2%; Pred. No. 3.2e-52;

Matches 304; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 21 caagctgaacatgagcaaaagccagagcagacagctgtaagtcaccaagctcagct 80  
 DB 58 CATGATGAGGAGGAGAACACACCATGAGACAGAAAGTAGGACTAATGAAGATGTTGAGCT 117  
 QY 81 gagaagaagccgttgagcagatggaagagatctcactcactcaactacatagcgaatca 140  
 DB 118 GAGGAAGGACCATGAGACCATGAGAGAGATCTCATCTCATCACTACATAGCCACCA 177  
 QY 141 cggcgaagcagcttggaactccctagcaaaagctgctgctgctgctgctgctgctgct 200  
 DB 178 TGGCGAGGAGTCTGGAAACACGCTGCTGATCGCGGACCTGAAGAGACTGGAAAGAG 237  
 QY 201 ttgtcgctccggttgctgaactatctgcagccagcgtcggagaggaacatactaac 260  
 DB 238 CTGCAGGCTTCTGTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 297  
 QY 261 tgaagcagctcctcctgatactggaactgacatgccaagctggaagcaagctgctgctgct 320  
 DB 298 AGAGGAGCAGCTTCTGATCATGGAATCTACAGACAGATGGGGTAACAGATGCTCAAAAT 357  
 QY 321 tgcgaagcactcctccggaagagctgacaaatgagataaagaactctggaagcactagaat 380  
 DB 358 TGCAGAGGAGCTCCAGAGAGACAGATTAACGAATCAAGAACTCTTGAAGAGACCGAGCT 417  
 QY 381 ccaaaagcactcaagcaagcagagcttctcctgctgctgctgctgctgctgctgctgct 440  
 DB 418 CCGAAGGAATTCACAAATGCGCAATCATCTGAGGGCCAAATCTCATGCTTTCGATGA 477  
 QY 441 agcaagcacaagccacatg 459  
 DB 478 AACTAGTCCCATGACAGATG 496

RESULT 5  
 AY026332 1343 bp mRNA linear PLN 30-JAN-2002  
 LOCUS AY026332

DEFINITION	ORYZA SATIVA MYB TRANSCRIPTION FACTOR JAMB mRNA, complete cds.
ACCESSION	AY026332
VERSION	AY026332.1
KEYWORDS	GI:13177577
SOURCE	
ORGANISM	Oryza sativa.
REFERENCE	Oryza sativa.
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
TITLE	1 (bases 1 to 1343)
REFERENCE	Lee, M.-W., Qi, M. and Yang, Y.
REFERENCE	A novel jasmonic acid-inducible rice myb gene associates with fungal infection and host cell death
REFERENCE	Mol. Plant Microbe Interact. 14 (4), 527-535 (2001)
JOURNAL	21304658
PUBMED	11310740
REFERENCE	2 (bases 1 to 1343)
AUTHORS	Lee, M.-W. and Yang, Y.
TITLE	Direct Submission
JOURNAL	Submitted (23-JAN-2001) Plant Pathology, University of Arkansas, 217 Plant Science Bldg., Fayetteville, AR 72701, USA
FEATURES	location/Qualifiers
source	1. 1343
	/organism="Oryza sativa"
	/cultivar="Drew"
	/db_xref="taxon:4530"
CDS	127..984
	/note="related to host cell death and defense responses; induced by jasmonic acid, wounding, or infection of rice blast fungus, but not by salicylic acid or abscisic acid"
	/codon_start=1
	/product="Myb transcription factor JAMB"
	/protein_id="AAK08983.1"
	/db_xref="GI:13177578"
	/translation="MEMVLTQTHHPVPGEDQEAELSAELRGPVTDEDTLLTNLYSHSGRMNARARAGCKRTKSCSCILRYNRYRPPVKRNPAAEDLLDHSRKGNSKTAQHLPGFTDNRIKRYKTRQKNAKQKANDGVNSRRDANKLMPRLAEIHAKAGAYDDSGDISNDLSCVSGTATYATVANCDSFMSYSSDYSODLKKIENLVHGDGKMNSEDDMEQVEDHFEWSTETLQPNPDQDQDLNGVGFSGLSLETLMLSDLEIMKQ"
BASE COUNT	367 a 298 c 337 g 341 t
ORIGIN	
Query Match	22.4% Score 192.8; DB 8; Length 1343;
Best Local Similarity	74.7% Pred. No. 1.3e-43;
Matches	242; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
OY	76 gagtggaagaaaaggcgctggagatggaagaaggatctcatctcctaacaactagacg 135
DB	208 GAGTCGCGGCGGCGGCGGCGCTGTGACCTGTGACGCGGAGCTCTACCTCATTAATTAATCTCT 267
OY	136 aatcagcggaaggagatgtgaaatcccttagccaaagatgtgtgtcttaaaacgtacggg 195
DB	268 GATCAGCGCGAGCGCGCTGTGAAACCTACCTGCGACGCGCGCGCTGTGAAGAGGACTGGG 327
OY	196 aagatgttgagctccggctgtgcttgaacatctcggagcccgagctccggagaggaactc 255
DB	328 AAGACTCTGCGGCTCCGCTGCGGTGAACATATCTCCGCGGAGTGAAGCCCGCAACTTC 387
OY	256 actactgaggaagctccctgatlcatgtgaactctatgcgaagttgggaaacagatgttct 315
DB	388 ACCCGAGAGGAGACACTCTCTCATCTTCGACCTCCACTCCCGATCGGCAACGATGCTCC 447
OY	316 aaatctgaagacatctcccggaagagctgacatcgaaataagaaagaaactctggagact 375
DB	448 AAGATGACCAACAATTTCGCTGGAGAGACGACCAACGATCAAACTACTGAGGAGAC 507
OY	376 agaatccaaagacatcaagcaa 399
DB	508 AGATGCAAAAAGCATGCCAACA 531

[illegible]

[illegible]

Query Match	19.5%	Score 167.8;	DB 8;	Length 1423;
Best Local Similarity	69.28;	Pred. No. 1.7e-36;		
Matches 229;	Conservative 0;	Mismatches 102;	Indels 0;	Gaps 0;

Oy	68	aagbltgcgaagltgaagaaaggscctggtgcgatggaagaggtctatctatcataact	127
Db	198	ACGAGATGAGACCTTAACAGAGAGGTCCGTGGATCGCTGAAGAGATTTTAAGCTCATGAAAT	257
Oy	128	acataagcgaatcaacgagcgaaaggcagltgtgaactcccaagccaaagctgcgtgtctaaac	187
Db	258	ACATTGCTACTAAATGGAGAAAGGTGCGTGAACCTCTTTTCGTGCGCGCGCTCCAA	317
Oy	188	gtaccgsgaagaagtttgcgcgtccggtgcgtgaaactcttcggaaccggaagctcgaag	247
Db	318	GCACCGGTAAAAAGCTGTAGACTAAGGTGGTTAAACTATCTCCGCTCGAAGTCCGCCGTG	377
Oy	248	gcaacatcactactgtgaagagcagctctctgtatcatgaaactgtcatgccaagtgggsaaca	307
Db	378	GAACATTCACCTTGAAGAACATCTTGATTCCTCGCAACTTCATCTCCGTTGGGGAAATA	437
Oy	308	gtgtgtctaaaatgtgcaagaacatcttcgcggaagagacttgcgaatgaataaaactctc	367
Db	438	GATGCTCAAAAATGCGCACAATATTTACCGGGACAGACGACAAACGATTAAGAACTACT	497
Oy	368	ggaggaactagaatccaaaagcacatcaaga	398
Db	498	GGAGAGCCGGGTGCAAAAAGCATCCGAACA	528

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
8	AP003140	AP003140	133242 bp	DNA	linear	PLN 22-MAY-2001	
		Oryza sativa genomic DNA, chromosome 1, BAC clone:OSJNBa0025P13.					
		AP003140					
		AP003140.2	GI:14164491				
		Oryza sativa (cultivar:Nipponbare) DNA, clone:OSJNBa0025P13.					
		Oryza sativa					

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (sites)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
*Oryza sativa* nipponbare(GA) genomic DNA, chromosome 1, BAC  
clone:OSUNBA0025F13  
Published Only in DataBase (2001) In press  
2 (bases 1 to 133242)  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Direct Submission  
Submitted (31-JAN-2001) Takuji Sasaki, National Institute of

**COMMENT**

Submitted 11-JAN-2001) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kamonada 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@ab.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel: +81-298-38-7441, Fax: +81-298-38-7468)  
On May 21, 2001 this sequence version replaced g1:12641874.  
Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN.0, BLASTX.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant Protein database, nr (<http://ncbi.nlm.nih.gov/blast/db/>) and the cDNA sequence database at RGP. Protein homologs of the coding regions were searched against NCBI Nonredundant Protein database with BLASTX2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DBJ accession no. and RGP clone ID.  
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level, such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.  
The orientation of the sequence is from -2M13 to M13rev of the BAC clone. This sequence of OSJNa0025P13 clone has an overlap with

p0702012 (DDBJ: AP002820) clone at the position 1 to 1534 of 5' end and with p0487H02 (DDBJ: AP002883) at the position 84175 to 133242 of 3' end. The sequence of this clone starts at the position 135799 of p0702012 and ends at the position 49068 of p0487H02. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://trp.dna.afrc.go.jp/Genomeseq.html>.

## FEATURES

## SOURCE

```

Location/Qualifiers
1..133242
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="OSJNBa0025p13"
complement(join(4221..4246,4864..5040,5244..5909,
6141..6189))
/gene="OSJNBa0025p13.1"
complement(join(4221..4246,4864..5040,5244..5909,
6141..6189))
/gene="OSJNBa0025p13.1"
/codon_start=1
/product="putative protein kinase"
/db_xref="GI:14164492"
/translation="MYLPLILVLAALVYISICRIGRRKKYKIHQGDIDQENM
FVYNFEGGNAVEITATEEMSEKTCGAGVSYVAALSKGRKAYKATENG
RLINEQFYEIEATMOIRHNITVKEGVCCTARDFIVTKYKGGNLTAIKSYSA
SELDKMRRLCADIQVAHALSYLHHDCSDPIVHRDVTKNILLDEFRACISDFGKSI
LDADGSGHTRLAGIKGLAPELAYTTVTEKCDVSECVVLELMLGSHGDFVSSIS
CPSKSTPMKDLDTLRPPAEVIALKIM"
join(9270..9380,9468..9560,9738..9797,9871..10143,
10245..10319,10435..10491,10609..10749,10829..10879,
10972..11067,11175..11258,11346..11448,11534..11616)
/gene="OSJNBa0025p13.2"
join(9270..9380,9468..9560,9738..9797,9871..10143,
10245..10319,10435..10491,10609..10749,10829..10879,
10972..11067,11175..11258,11346..11448,11534..11616)
/gene="OSJNBa0025p13.2"
/notes="contains ESTs C72171(E1142), AU093893(E1142)"
/codon_start=1
/product="putative shaggy-related protein kinase"
/protein_id="BAB55743.1"
/db_xref="GI:14164493"
/translation="MTSGVLVDSSSGEPSTSGADRLTDDISEMSIRKEEAVV
SGSMDIGHTVTVGNGSQPKOTISTYIAERVAGSGFVVAQACLEGEEVAARK
VLQDARKKNEELDTMOVLDPHNAALNHYCSPTAKELVLTNLEVEPTVIRVIRH
YNNKSRMPRLIVKLYMYQICRALAYHNCVCYCHRDIKPONTLVNPHNQLCDG
SAKVLVGEPNISYICSRYYRABELFAGAREYTAIDVMSAGCVLALLGQVPGD
SGVQLVEITIKVLEPTREEIKHNPNYTEFEPKPKAIKHPMKIENHRMSEAVDLVS
RILOYSPLHSCALEVLIHPPFDELRRPNRLNGRLPLPLNFKPRELKGASMEFLV
KLVQNAKKQCAFLEGT"
join(15137..15820,15928..15963,16058..16849)
/gene="OSJNBa0025p13.3"
join(15137..15820,15928..15963,16058..16849)
/gene="OSJNBa0025p13.3"
/codon_start=1
/product="putative polygalacturonase-like protein"
/protein_id="BAB55744.1"
/db_xref="GI:14164494"
/translation="MELAAAGRTAAILALLAFASSFISAADGARSARHHNAKNAK
MAHPQDAPGARPARHAPRHGAPRHGAPRHGAPRHGAPRHGAPRHGAPRHGAPRH
PAGCAATVIVIDVDFGAAGDVDDDTALCTAMDGTACADGACVLAAGRSRLIHT
VTFPGCGSTTLOYDGTIVAPSEPAITPANNKRMILVTRADYSLVGAGLIDGCKG
WMDLPLPKRGKGNTHGPCDSVAMRFATISNNVTVKGVNSPEHFRFNCGVARD
GLSDPLASPNIDGIVHENTSDVLIINTVNGDVCISGAGTLNHNIEINCPGPH
GISISGLKAGTKACVANTVRNVAIRHSDNGVAIKWOGSSGSVAEENRMAV
RNPILIDQYCTKSCENETIAVEVNGSVSGIRGTVDGRPIHGGCSAVYCTNIT
LSVDELLPAGSDTVDEBECNVYGNATTPVPPSCSSGEPNTEKKDQCY"
complement(join(21118..22066,22246..22397,23530..23691))
/gene="OSJNBa0025p13.4"
complement(join(21118..22066,22246..22397,23530..23691))
/gene="OSJNBa0025p13.4"
/notes="hypothetical protein"

```

```

/codon_start=1
/protein_id="BAB55745.1"
/db_xref="GI:14164495"
/translation="MRAAEQLRVAQAARAPGARWTHRGARDYRGGCGDRSARP
ATEAGCARVYTKNGTQTRSPYSGVYKRJUEDEKRTTAAVHQSKEKRLCPRAIT
LGLRAGRAPVPPHAPAPRTVEVASSSTTSASAKVNSAATNLSHRLTSRO
ACILHLRPLSRACGRCHRLHVSYSRSPTPPNNMALADTTYCHALGCKPBEAMPP
PTTVVTSPPSPPTQLPLRHCHRLHVSYSRSPTPPNNMALADTTYCHALGCKPBEAMPP
HCKPAPPTQPPQRRRLALCRDRHGRKTISSAMVRSPPASPLDPAVAVLTSI
PAISNPADAPSKRSSMPARTPSLSPPPPPSPSSAPATTCNRRHPTTFPRR
HHHQLPLPHPTDPPPAFPDPDATAARI"
complement(join(27109..27305,28392..28419))
/gene="OSJNBa0025p13.5"
complement(join(27109..27305,28392..28419))
/gene="OSJNBa0025p13.5"
/codon_start=1
/notes="hypothetical protein"
/protein_id="BAB55746.1"
/db_xref="GI:14164496"
/translation="MDVVIIGKTKKQTANQKKANAMHSQDITSSIGNFDIDKK
DNLNRSYTVAFTEYRSIVANRLALIK"
complement(join(31012..31020,31243..31294,31358..31478,
32907..33108,34245..34337))
/gene="OSJNBa0025p13.6"
complement(join(31012..31020,31243..31294,31358..31478,
32907..33108,34245..34337))
/gene="OSJNBa0025p13.6"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB55747.1"
/db_xref="GI:14164497"
/translation="MHERVLSYNDKEITGFHNPOLCHMGKQKERTHPLPVRNR
MPSDKFAENMRANOSLSFDVSHVHFVIVLLASVLSKRSKRSPTDQGRVSR
TSVATSCSDCARIPSVYVNVILPILSSGMEDRLVCPPLFVPIGFGOKPVPNW"
41044..44517
/gene="OSJNBa0025p13.7"
41044..44517
/gene="OSJNBa0025p13.7"
/codon_start=1
/product="putative GAG-POL precursor"
/protein_id="BAB55748.1"
/db_xref="GI:14164498"
/translation="MPCVPRVYEHRLAVPGARPVROKVRQADPEROAFIREVAVRL
LEAGFIREYIHEMLVNPVIVKANKKLMCIDYDLMACXKDYPLSLRIDQYDST
AGCDLCTFISGTHQIMNAREDEKTAFTIPIGTCITTPFGIKRNGPFOKTR
ISGSQIGNVEAYVVDLVKTRNOFTLLSDLAETFEFSLSARIKILMPDKCVGPVAG
KILGFLVSARGIEANPEKIRAIERMPSPKLDVQCVTCMALESFISRLGEKALPL
FKLKSGSEPTWEEAREFTOLKAVLSPPVLAPEDEPLLYLAAPQVSAALV
VERDENPISANHRPASTRPGEGGEAEPPGPPRPAAGAPRACPTVGAADPO
DGGGATAGRPRLSPSPPEVNGEAEAPRGSLDEEPPGGAARGEEDRPKRVQRPVF
VSEALDATTRPQAKMLAIVASRKLRHFQANRYVYVSTYPLGOLINRECTOR
VSMVALESEDFIREPRHAIISQALJTDVAEMTPAPEVSAPEASSGSQLPHTAVM
VWOFDLSLSOGAGVTLTSPSGDLRLVLDPRATNNMKEIGELRLHVPRLNAGA
IRLVLVIGDSQIVANQVYKEYKESCDPOMAVYOVARMRHRFDGIELRHVPRLNAGA
DELSRLASSRAQTPGAFEEERLAOPARADPLGETIDAPRRPRTIGVQSGEGSAPS
SPRLTAWTEIOTIYLTADKTLPEDRGSESVIRIYKRYVEGTLVRRANGVLLCIP
RQGVFLADLVHEGEGASASRTLYGKAFRGGEYVPTLNNAVLVRCRCQCHAR
QTHQPAQLOTLPLSWPFAVWCUDILGPRRABGGEYQVAVVDFTPKPEATPIKI
DKHSALKEIRGITARFQVNPRIITKNGSTQTEPLVLEGDCKDGIKICFASPRNSNG
OVERAAEELIKGKTFTFNIILKKNQSTIEKLPVLAAMARTKPSATGTFSTFLVGA
EVALPELIRSRVTVCEADODLROLDLYLEERRRALLAARYOOSLRRHOR
HVARSELCDVDLVLRVQTRAGLSKSPHMBEPRIVYICVPRRGSVRLATGCGTELPNP
WNEHILRRYPP"
complement(join(45140..45474,45602..46322,46382..47023,
47154..48413))
/gene="OSJNBa0025p13.8"
complement(join(45140..45474,45602..46322,46382..47023,
47154..48413))
/gene="OSJNBa0025p13.8"
/notes="hypothetical protein
similar to Oryza sativa chromosome 3, OSJNBa0067E01.7"
/codon_start=1
/protein_id="BAB55749.1"

```

```

/db_xref="GI:14164499"
/translation="MARGSAPLIDGSLPPSRIVSEROGALPPRPMESATGTEVVALG
KGRPAHYGWSVFLSPFAMAGLIVPEPSTFEMDILEFYDLOAHILTPAINTLAIPEAH
LCENMTIRVRPSRLRFWRFTTVVPVPPPAVGGCTYQPRGVYINRIIPCVLKKKWDIMK
SMWTFPLGEARLRPSOPPAQSSMWTAVDVGCTVADLDRGLASQSLGTAMVF
GYLLRRRLAFLORARAGMEYTPEDPMYTHOGKMDPDPEPSVYVORVNLSSVEA
SLTPGVPLICDPERAKITLIMQVVGASGERAPHDGTGSRGEGSTPGGASG
PDDGGGSRPADGARKRKQEGTPPPGCGAGAVASRSRREGAVPPPOPGGERKK
RLKKMGIVPCRGNSLQGTTEPHLAKRDPYAPPLPPPPPHSKSGSRSEAEYTAE
ARRRADREADRLREAEARAEAEVARYROAEARBEARERAESESRVABAAARAO
AACQTPDSTPEATTTSKATHEAEACAPLGPSPSDAMDEPDSGAPPSGTSIGPSR
AAPTRGLSPRSPAPLSEAPFLQALAAANSTVLDGLSNOMDALAEARLDAAMARV

```

Query Match 19.3%; Score 166.4; DB 8; Length 133242;  
 Best Local Similarity 68.5%; Pred No. 9.8e-36;  
 Matches 220; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 63 gtcccaagatgtcgaagtgagaaagggccgtgagacatggaagagatctcatc 122  
 DB 120124 GGGCGAGAGAGAGAGATGTTGGAGAGGGCCGTGGACGGTAGAGAGAGCGGTGGC 120183  
 QY 123 caactacataggaatcaaggcggaagggcagttggaactccctcctgaagctgctgtct 182  
 DB 120184 CGGCTACGTGGCGCGCAACGCGAGGGCGGTGCAACGACCTGGCGCGCGCGCGGCGCT 120243  
 QY 183 aaagctacgggaaggaagtgctgcgttcggtgagactatctgcgaccgagctcg 242  
 DB 120244 GAGCGCGACCGGGAGAGAGCTCCGCTCCGCTGCTCACTACCTCCGCCCGACGCTGC 120303  
 QY 243 gagaaggaacaatcactactgaggaagcagctcgtatcatggaactgcatgccaagtggg 302  
 DB 120304 CCGCGGCGACTTCAACCCCGCAGAGAGCTGCTCATCTCTGACCTCCGACTTCCGCTGGGG 120363  
 QY 303 aaacaggtgtctcaaatgcaagcatctctcccggaagagctgacaaatgagataagaa 362  
 DB 120364 GAACCGCTGTCCAGGATCCCGACGACATCGCGGGAGCAGCAACGACATCAAGAA 120423  
 QY 363 ctctcgagagactagaatcccaaacacatcaagca 398  
 DB 120424 CTACTGGCGGACCAAGGCTGCAAGAACGCGCAAGCA 120459

RESULT 9  
 LOCUS AP002883 145491 bp DNA linear PLN 15-SEP-2001  
 DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone: P0487H02.  
 ACCESSION AP002883  
 VERSION AP002883.2 GI:15623781  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa (cultivar: Nipponbare) DNA, clone: P0487H02.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 145491)  
 Sasaki,T., Matsumoto,T. and Yamamoto,K.  
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone: P0487H02  
 Published Only in Database (2000) In press  
 2 (bases 1 to 145491)  
 Sasaki,T., Matsumoto,T. and Yamamoto,K.  
 Direct Submission  
 Submitted (11-OCT-2000) Takuji Sasaki, National Institute of  
 Agrobiological Sciences, Rice Genome Research Program; Kannonndai  
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 (E-mail:tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,  
 Tel:81-298-38-7441, Fax:81-298-38-7468)  
 On Sep 14, 2001 this sequence version replaced gi:10800078.  
 Genes were predicted from the integrated results of the following:  
 GENSCAN1.0, BLASTX2.0, BLASTY2.0 as well as SplicePredictor  
 (October 1998 version). The genomic sequence was searched against  
 NCBI Nonredundant protein database, nr  
 (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at

## FEATURES

source

RGp. Protein homologies of the coding regions were searched against  
 NCBI Nonredundant protein database using BLAST2.0. ESTs represent  
 the identified cDNA sequences with BLAST2.0 with the  
 corresponding DBJ accession no. and RCP clone ID.  
 A gene with identity or significant homology to a protein is  
 classified based on the protein name to indicate the homology level  
 such as same name, 'putative' and '-like protein'. A gene without  
 significant homology to any protein but with EST homology (covering  
 almost the entire length of partial sequence) is classified as an  
 'unknown' protein. A gene predicted with a gene prediction program  
 is classified as a 'hypothetical' protein.  
 The orientation of the sequence is from SP6 to T7 of the PAC clone.  
 This sequence of P0487H02 clone has an overlap with P0682B08 (DBJ:  
 AP003578) clone at the position 73,961 to 145,491 of 3' end.  
 Detailed information on overlap and assembly quality together with  
 annotation of this entry is available at  
 http://rgp.dna.affrc.go.jp/Genomeseq.html.

Location/Qualifiers

```

1..145491
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0487H02"
complement(join(6..1328,2016..2113,2149..2286,2521..2676,
2722..2784,4696..4740,5045..5186,5198..5407))
/gene="P0487H02.1"
complement(join(6..1328,2016..2113,2149..2286,2521..2676,
2722..2784,4696..4740,5045..5186,5198..5407))
/gene="P0487H02.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB67842.1"
/db_xref="GI:15623782"
/translation="MQSKHGRSTFGTSQGLARCSVSKDEKYLCVNSCNPHPQTTT
HKCKRPSANARPNRNDARRGRTIPKRLDRDITRPSQKGLCTGTPLRSSQNHRTI
ITSYDYVYIVITGCLRPFVPSLSHPCSDLAGRCAGDNGVEMQHILAGGR
GEAGGGNDGTGAAALAPLPLPSQIMPEEGGRLAAVATMSGDVYPHHMETSTSD
DRKSRRTGOKLAHYLPPHMGSPSKYRRSEITLPLRKEELOREDAECGELHRTI
AADQREPRPNSSESLTTPRSIPLSLPMDQRLRRMAADDDGYLEYARRITPEITDR
RHMGRVNSWRVALVKLAPAPPPPLPLALPESDDGLPATVSCVLSGCRHASVLO
GARGARVFGSYDGMFLPVGGOACQALMLKINGDTGLDPLNAPVNSVNPNRDR
MAIVAAATLSQPTGEGCIAGIIESPPNIVAGVTRSIJAFWRMGDOVLYLVLAEE
DNPLMRLEEVEDLCHGAPFHTFAEVLVLAECSEPVPYRDSVSLVPANMFLRVHD
ENETYLARIYVSGSKMLMYRLASGRQRTTSAPRVQKKPTGDEDSQSRSH
FEIYSELDEDRMLFVGLCSRSYKAGDRTPEMERGVTFDDPSTIHQITGDAPK
PPLYCSNDGKMSRAPTPDQGOVERCFPERGPSIHSPVWILP"
join(7576..7805,8020..8053)
/gene="P0487H02.2"
join(7576..7805,8020..8053)
/gene="P0487H02.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB67843.1"
/db_xref="GI:15623783"
/translation="MNPRIINLVLHVHGVOKFLVNFSSCRENNPEKISPRISQAA
AYTRVARNRNRHRAACGCMERGNRRMASSNLTATHTART"
join(8658..8923,8973..9006)
/gene="P0487H02.3"
join(8658..8923,8973..9006)
/gene="P0487H02.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB67844.1"
/db_xref="GI:15623784"
/translation="MRATVGQATTHDQASHARTAAVAAGVARDTPISLOVHVCMSG
AMVETNTEPEPAOGSKKRGVYLISKFTCIHOTPEYLSPLDPSSPYQIISRVA"
complement(join(10181..10287,11338..11896))
/gene="P0487H02.4"
complement(join(10181..10287,11338..11896))
/gene="P0487H02.4"
/note="hypothetical protein"
/codon_start=1

```



[illegible]



PUBMED	11597504
REFERENCE	2 (bases 1 to 1140)
AUTHORS	Stracke,R. and Weisshaar,B.
TITLE	Direct Submission
JOURNAL	Submitted (10-JAN-2001) Dept. Plant Breeding and Yield Physiology, Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln D-50829, Germany
REFERENCE	3 (bases 1 to 1140)
AUTHORS	Stracke,R. and Weisshaar,B.
TITLE	Direct Submission
JOURNAL	Submitted (29-AUG-2001) Dept. Plant Breeding and Yield Physiology, Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln D-50829, Germany
REMARK	Sequence update by Submitter
COMMENT	On Aug 30, 2001 this sequence version replaced gi:13430155.
FEATURES	
source	1..1140 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /note="ecotype: Columbia"
gene	1..1140 /gene="MYB16"
5'UTR	1..129 /gene="MYB16"
CDS	130..981 /gene="MYB16" /note="Atlg25340; R2R3-MYB family member" /codon_start=1 /product="putative transcription factor MYB16" /db_xref="GI:15375295" /protein_id="AK25748.2" /translation="MSNITKKKNGNEGABORQKPWLTLEDLLTNYISHGGRNN LAASSGIKRAGKSCRIWLMLYLDPDKIRGLTFPOEOLLILEHSKGNRMSKSYLU PGRPNDINKWNTFRVROAROLINDSNHKEIVVSFMFRLINELKDNRSTNNJRK AMADDELGPLRDOKDLGFNMDCSTSMEDLKTSOPMDFSDLETLMELEGSGSS OCSEEVYSPECLLEEYRVAVMGSSDISALHDCHVAADSKIEDVTOTLMRMDDIMQPR NEIYAHFN 982..1140 /gene="MYB16"
BASE COUNT	397 a 202 c 222 g 319 t
ORIGIN	
Query Match	18.8%; Score 162.2; DB 8; Length 1140;
Best Local Similarity	67.8%; Pred. No. 6,2e+35;
Matches 227; Conservative	0; Mismatches 108; Indels 0; Gaps 0;
OY	55 agtggtaagtcccaagtatgtcaggctggaagaaggcgctggaacatagaaagatctc 114   Db 157 AATGGAATGAAAGACGGGTGCACAGCAGAGAAGGCGCTTCGACACTCGAGAACACT 216
OY	115 atcccatcaactacatattgcgatcacggcgagcgatgttgaaactccctagccaaagt 174   Db 217 CTTTCACCACAATTACATTTCCCATTAACGGTGAAAGGCCGATGGAATCTGCTCAATCT 276
OY	175 gctgtgtctaacaagctaccggaagaagttgtcggctcggctgacctatctgcgaccc 234   Db 277 TCTGGGCTTAAAGACAGACGAAAAGTTGTACATTGAGATGGTTGAATTTACTTTAAACC 336
OY	235 gaagtcggagagggcaacatcaactctgaagagcgactctgatatgaaactgtacgc 294   Db 337 GACATAAAGCCGGGGAATCTCACTCCTCAAGAACAACTTTTAATCCTTAGCTCATTTCT 396
OY	295 aagttgggaagaagctgtgtctaaaattgcaaacgactctccggagaagagctgacaatg 354   Db 397 AAATGGGGTAATAGGTGGGCTCAAAAATTTTCAGAGTATTTTACACAGAAAGACAGACAAACGAT 456
OY	355 ataagaactcttgagaggaactagaatccaaaaagca 389   Db 457 ATCAAAGAACTACTGGAAGAACTGAGAGTCCGAAACA 491
RESULT	11

	AY008377		931 bp	mRNA	linear	PLN 30-AUG-2001
LOCUS	AY008377					
DEFINITION	Arabidopsis thaliana putative transcription factor MYB112 mRNA,					
ACCESSION	complete cds.					
VERSION	AY008377					
KEYWORDS	AY008377.2 GI:15375307					
SOURCE	'					
ORGANISM	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 931) Stracke,R., Weber,M. and Weisshaar,B. The R2R3-MYB gene family in Arabidopsis thaliana Curr. Opin. Plant Biol. 4 (5), 447-456 (2001)					
REFERENCE	AUTHORS TITLE JOURNAL MEDLINE PUBMED					
REFERENCE	PUBMED AUTHORS TITLE JOURNAL REFERENCE					
JOURNAL	2 (bases 1 to 931) Stracke,R. and Weisshaar,B. Direct Submission Submitted (29-SEP-2000) Dept. Biochemistry, Max-Planck-Institut fuer Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln D-50829, Germany Zuchtforschung, Carl von Linne Weg 10, Koeln D-50829, Germany Sequence update by Submitter On Aug 30, 2001 this sequence version replaced gi:11641119. Location/Qualifiers 1..931 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /note="ecotype: Columbia"					
CDS	misc_feature CDS misc_feature					
5'UTR	6 /note-"alternative transcription start site" 84..815 /note-"Atlg48000; R2R3-MYB family member" /codon_start=1 /product=""putative transcription factor MYB112." /protein_id="AAG38379.2" /db_xref="GI:15375308"					
3'UTR	/translation="MNISRTFPAACKTLINKKEVEYEKMEIEIRGPVTVEDMKL LVSYSILSGEGRWMSLSASGLNRKSCRLRMNLNYLRPDIRDSDISOEQFIILELH SRWSRKMSIAQHLFGRTDNELKNWRTRVOAKLLKCVDNSKFQDTIKHMMPRI IERLATOSVFOTSNHYSPENSVAATTSSTSSSEAVRSFEYGCDGVFFCLDMTNC GYMFMGGDPFTFLCSFDLELNKLMLQ" 816..931					
polyA.site	897					
BASE COUNT	317 A 183 C 194 G 237 T					
ORIGIN						
Query Match	Best Local Similarity	18.0%;	Score 155.4;	DB 8;	Length 931:	
MATCHES	Matches 234; Conservative	0;	Pred. No.5,le-33;	Indels 0;	Gaps 0;	
D	ctgaacatgtagacaagaagcccgaccagcgcagtcgttgaaagtcccagaagctcgagggtgaga	84				
Oy	123 CTATTATAATCATTAAGACAAGAAGTCGAAGAGCGAGAAAAAAGATGTGAATTAGAATAAG	182				
D	aagggccgttggacgatggaaaagattcatcaactccaatactaatacgaatacacgyc	144				
Oy	183 AGAGGCCATGCAGCTGTGGAAAGAACATCAAGACTCCTGATTTCTTGCCAGCGT	242				
D	gaagcagtttgtgaacctcectagccaaagtctgctgtcttaaacgftacogggagaaggttgt	204				
Oy	243 GAAGGAAGATGGAACTCCCTCTCTCTCTCTGTCGTGAGCATGAATAGAAACGGGAAAAGTTGC	302				

Oy	205	cgagcccgatgagcctaactatctgagaccacgcgcgcggaagaggaactaactactag	264
Db	303	AGATTGGCGGTGAATTATCTCCGGCCGGATATCCCGCGGAGACATATCCTTCAA	362
Oy	265	gagcagcctccatcatggaactgcatagccaaagtgggaagaacagatggtcctaaattgca	324
Db	363	GAAACAATTATATCATCTTGAACATCATTCGTGTTGGGAAATCGGTGGCAAGAATTCT	422
Oy	325	aagcactctcccggaagagactgacaaatgagataaagaactctggaagactagaatcaca	384
Db	423	CACACTTTTACCGGAGAACACATTAACGAGATAAAGAAATTATTGGAGAACACGCTTCAA	482
Oy	385	aagca	389
Db	483	AAGCA	487

AF034133	LOCUS	AF034133	1033 bp	mRNA	linear	PLN 02-MAR-1998
DEFINITION	Gossypium hirsutum MYB-like DNA-binding domain protein (Cmy-N)					
ACCESSION	AF034133					
VERSION	AF034133.1	GI:2921337				
KEYWORDS						
SOURCE	upland cotton.					
ORGANISM	Gossypium hirsutum					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.					
REFERENCE	1 (bases 1 to 1033)					
AUTHORS	Loguercio, L., Zhang, J. and Wilkins, T. A.					
TITLE	Structure and expression of six classes of myb-domain genes in					
	allotetraploid cotton ( <i>Gossypium hirsutum</i> L.)					
JOURNAL	unpublished					
REFERENCE	2 (bases 1 to 1033)					
AUTHORS	Loguercio, L., Zhang, J. and Wilkins, T. A.					
TITLE	Direct Submission					
JOURNAL	Submitted (13-NOV-1997) Agronomy & Range Science, University of					
	California, One Shields Ave., Davis, CA 95616-8515, USA					
FEATURES	Location/Qualifiers					
source	1..1033					

gene	CDS	repeat_region	repeat_region	BASE COUNT	ORIGIN
gene	1..1033	repeat_region	repeat_region	308 a	
/gene="Cmy-N"	/note="MYB-domain gene N"	106..987	106..987	223 C	
/gene="Cmy-N"	/function="putative MYB-like transcription factor"	Number L04497	Number L04497	253 g	
/codon_start=1	/product="MYB-like DNA-binding domain protein"			249 t	
/protein_id="AAC04719.1"	/db_xref="GI:2921338"				
/translation="MPEREVSYLKLQEKRRMRRLRKGPMTEEDSMRLAYVNIHGEGR					
WNAVARLSGLRRGTGSCRLRLNLYMRPEIKRGNLSLEPLLILHLRWGMRKING					
HLPGHTDNEIKNYMTRKYQAKQAKLQNDYNSKQKRDNRVYVIRLRYIRASSEPS					
SQPTSTNTFTYNDRLSNISSSQSTAAANSQVQVDPSTLPLSTGSSSDLTQVSVAS					
DLTQCYNQPSLHLKLGKLGLEKGAATWGVDEFOATEHSNGMLVGGGSSMDTW					
NEENWFLQOOLHDPDGI"					
163..318					
/note="putative MYB DNA-binding domain repeat R2"					
319..474					
/note="putative MYB DNA-binding domain repeat R3"					

Query Match	17.6%	Score 151.4	DB 8	length 1033
Best Local Similarity	67.0%	Pred. No. 7.1e-32		
Matches 215	Conservative 0	Mismatches 106	Indels 0	Gaps 0

QY	78	gctgtaaaaggccgctgacatctgaagagatctccctcaactcaactcatagcgaa	137
Dp	162	gctaaagaaaggtccatgacatgaaagaaagacatcgtttagccgctatgtcaatatt	221
QY	138	tcaagcgaaagcgatctggaatccctctgccaagctctgctcttaaaagctaccggtgaa	197
Dp	222	ccacggcccaaggtccctctgaaacgctgctgctcctcttatcagattgaacaacccggtaa	281
QY	198	gagtcgtcgtctccgctgctgtaactatctgcagccgacgtcccgagaggaacacatcac	257
Dp	282	aagctgcagattgaatgactgtgaacttatatgcgaccacaataatcaacaggaatacatgag	341
QY	288	tactgaaggaagaagctccctgatactgaatctcatgccaagctggggaataaagatgctcaa	317
Dp	342	ccttcgaagaagcagctattgattcttgaaactcattcttcgcttggggaacagatgctcga	401
QY	318	aatttcaaaagcatctctcccgaaaggaactgacatctaga taaagaaactctctgaggaactg	377
Dp	402	aatttcacaacactcgtccgtgaaagaaacagacatcaatcaatataatgttgagaaacag	461
QY	378	aatccaaagacatcatcaagca	398
Dp	462	agttccagaaagcagcccaagca	482

```

RESULT 13
LOCUS AF427146 1513 bp mRNA linear PLN 29-OCT-2001
DEFINITION Zea mays myb-like transcription factor (P) mRNA, complete cds.
ACCESSION AF427146
VERSION AF427146.1 GI:16507119
KEYWORDS
SOURCE
ORGANISM
Zea mays.
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1513)
Hartmann,U., Sagasser,M. and Weisshaar,B.
Separation of cis-acting elements required for light-responsive and
tissue-specific expression of co-regulated phenylpropanoid
biosynthesis gene
Unpublished
2 (bases 1 to 1513)
Hartmann,U., Sagasser,M. and Weisshaar,B.
Direct Submission
Submitted (01-OCT-2001) Salaminl. Max-Planck-Institut fuer
Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, NRW 50829,
Germany
FEATURES
source
location/Qualifiers
1..1513

```

```

/strain="P-rr-4B2"
/db_xref="taxon:4577"
1..1513
/gene="p"
179..1186
/gene="p"
/note="DNA-binding domain"
/codon_start=1
/product="myb-like transcription factor"
/protein_id="AA124047.1"
/db_xref="GI:16507120"
/translacion="MGRTPCCEKVGKRGWTAEDQLANYIAHEGSGNRSIPKNA
GLGKSGCRLRWINLRADVAPGNIISKEEDDILIKLHATIGNRMSLIASHLPRTON
EIKRYWNSHLSROIHTRRKYTAGPDOTALIDMSKLOSADRRGCGTPGCPAASAS
RTKQADADQPCGEAKGPAAASPRHSDVVPNGQNPSSGSGTGAEEVDPSSDDAS
GPTWLEPTELGLDGLWGEADSEMDALMPIGPGGHSALELGIAVGCERAOVDLEDMW

```





[illegible]



OX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Bann J., Carrincci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shino P., Yamada K., Shinozaki K.,  
 RA Becker J., Theologis A., Davis R.W.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF042894; AK68834.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 293 AA; 32221 MW; A12532EE45795A CRC64;

Query Match 38.5%; Score 574.5; DB 10; Length 293;  
 Best Local Similarity 47.0%; Pred. No. 4e-43;  
 Matches 133; Conservative 17; Mismatches 62; Indels 71; Gaps 6;

OY 5 AETRYDSFETTSGGSSVDLGMALPGLAPNLSSASVSASASASAKKIRKPYITTKSRSS 64  
 DB 14 AETSTDAETTT-----TATTEGAPEKKVKKATYITTKSRSS 50  
 OY 65 WSGQENDKFLKALQFLDRDKKIEAFVSGKTYIQRSHAKYFLVQKNGTREHVPPEPP 124  
 DB 51 WTEGHDKFLKALQFLDRDKKIEAFVSGKTYIQRSHAKYFLVQKNGTLAHVPPPP 110  
 OY 125 KRKASHPYPOKASKNPVSGVSTAPPTATQDSCGYTPRAESSITLKSGSSCPTVSSM 184  
 DB 111 KRKAHPYPOKASKNPVSGVSTAPPTATQDSCGYTPRAESSITLKSGSSCPTVSSM 164  
 OY 185 VHHHTPSIDASFVEKDDGSPGIEGNNCS-----GSTESSPTWPCSEIPEKVK----- 236  
 DB 165 I-----PPEDELDTLCAEYDVGSNDMISETSASASGIGSSSRILSD 206  
 OY 237 -----PDFSQYKFTGVSVPDPTDHLKKLKE 263  
 DB 207 TKGRLAKQAPSMHGLPDAEYVNFISGVDPDPSKGRMKLKE 249

RESULT 5  
 O91Z01 PRELIMINARY; PRT; 307 AA.  
 AC O91Z01;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE HYPOTHEICAL 33.9 KDA PROTEIN.  
 GN F9G14.150.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Ardiles M., Buyschaert C., Dasseville R.,  
 RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van den Daele H.,  
 RA Villarroel R., Glendon J., Van Montagu M., Bancroft I., Mewes H.W.,  
 RA Rudd S., Lemcke K., Mayer K.F.X.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL162973; CAB86038.1; -;  
 DR InterPro: IPR001005; MYB\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 1.  
 DR SMART: SM00395; SANT; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 307 AA; 33896 MW; 0F1C6D375A62B050 CRC64;

Query Match 38.0%; Score 567.5; DB 10; Length 307;  
 Best Local Similarity 44.8%; Pred. No. 1.8e-42;  
 Matches 133; Conservative 17; Mismatches 62; Indels 85; Gaps 6;

OY 5 AETRYDSFETTSGGSSVDLGMALPGLAPNLSSASVSASASAKKIRKPYITTKSRSS 64  
 DB 14 AETSTDAETTT-----TATTEGAPEKKVKKATYITTKSRSS 50  
 OY 65 WSGQENDKFLKALQFLDRDKKIEAFVSGKTYIQRSHAKYFLVQKNGTREHVPPEPP 124  
 DB 51 WTEGHDKFLKALQFLDRDKKIEAFVSGKTYIQRSHAKYFLVQKNGTLAHVPPPP 110  
 OY 125 KRKASHPYPOKASKNPVSGVSTAPPTATQDSCGYTPRAESSITLKSGSSCPTVSSM 184  
 DB 111 KRKAHPYPOKASKNPVSGVSTAPPTATQDSCGYTPRAESSITLKSGSSCPTVSSM 164  
 OY 185 VHHHTPSIDASFVEKDDGSPGIEGNNCS-----GSTESSPTWPCSEIPEKVK----- 236  
 DB 165 I-----PPEDELDTLCAEYDVGSNDMISETSASASGIGSSSRILSD 206  
 OY 227 PCSEIPEKVK-----PDFSQYKFTGVSVPDPTDHLKKLKE 263  
 DB 207 SASGIGSSSRILSDSKGLRLAKQAPSMHGLPDAEYVNFISGVDPDPSKGRMKLKE 263

RESULT 6  
 O04605 PRELIMINARY; PRT; 213 AA.  
 AC O04605;  
 DT 01-JUL-1997 (TREMblrel. 04, Created)  
 DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE A\_IG002N01.20 PROTEIN.  
 GN A\_IG002N01.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Scheel P., Maggi L.;  
 RT "The sequence of A. thaliana IG002N01."  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Wash-U;  
 RT "The A. thaliana Genome Sequencing Project";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Waterston R.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF007269; AAB61027.1; -;  
 DR InterPro: IPR001005; MYB\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 1.  
 DR SMART: PS00395; SANT; 1.  
 DR PROSITE: PS00900; MYR.3; 1.  
 SQ SEQUENCE 213 AA; 23616 MW; 8C76CCEB9AB87A68 CRC64;

Query Match 33.7%; Score 502.5; DB 10; Length 213;  
 Best Local Similarity 56.4%; Pred. No. 6.7e-37;  
 Matches 106; Conservative 20; Mismatches 45; Indels 17; Gaps 5;

OY 22 DLVGMALP-----GLAPNLSSASVSASASAKKIRKPYITTKSRSSBDGDKFLK 76  
 DB 15 DSSNMSLPSSDGGFSIPATGRS-TVSFSDPTTKIRKPYITTKSRSSNTDDEHDKFLK 73  
 OY 77 LQFLDRDKKIEAFVSGKTYIQRSHAKYFLVQKNGTREHVPPEPPPKRASHPYPOKA 136



```

Db      74 LHLFEDRMKKLEAFVGSKTVQVIRSHAQKVELKVGKSGANHELPKPPKRAASHIPYITKA 133
QY      137 SKNPVYSQGVSTAFPTAT--QLDGTYIPRAESSILTKSSSCPTVYSSWVHNTI----P 190
Db      134 PKNV-----ATSTLSPSSSTLLPGLPGYLLSSDSKSLMGONAVCASTSSSWHNESTLNPKP 188
QY      191 SIDA5FEV 198
Db      189 VIEGSEFTD 196

RESULT 7
Q9SF43 PRELIMINARY; PRT; 125 AA.
AC Q9SF43;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE MYB-RELATED PROTEIN.
GN F11F8_19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eusteroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maiti R., Romling C.M., Koo H., Fujii C.Y., Utechtack T.R.,
RA Bartschend M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 3 BAC F11F8 genomic sequence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC016661; AAF23291.1; -.
DR InterPro; IPR001005; MYB_DNA_bind.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00395; SANT; 1.
DR PROSITE; PS50090; MYB_3; 1.
SQ SEQUENCE 125 AA: 14063 MW: 84924097F3F726EB CRC64;
```

[illegible]

RA Neut P., De Clercq R., Compliciers J., Denais P., Villarroel R.,  
RA Rouze P., Van Montagu M.;  
RT "Sequence analysis of a 40-kb Arabidopsis thaliana genomic region  
RT located at the top of chromosome I.",  
RL Gene: 215:11-17(1998).  
DR EMBL: Y12776: CAA73305.1; -.  
DR InterPro: IPR001005: MYD\_DNA\_bind.  
DR Pfam: PF00249: myb\_DNA-binding; 1.  
DR SMART: SM00393: SANT; 1.  
DR PROSITE: PS50090: MYB\_3; 1.  
SQ SEQUENCE 162 AA: 18929 MW: 34D5D44FLD796FAF CRC64;

```

Query Match          30.9%; Score 460.5; DB 10; Length 162;
Best Local Similarity 73.0%; Pred. No. 2.6e-33;
Matches 89; Conservative 12; Mismatches 12; Indels 9; Gaps 2.

QY 26 MALPG--LAPNLSSASVSASAS-----EDSAKKIRKPYITTSRSRSSEOEHDFLEA 76
    |.|||.||.::||.::||.|||||||.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
Db 16 MSLPGENTLPHATTTPVSISSNRMSFEPDPTKKVRPYITTSRSRMTEGEHDFLEA 75

QY 77 LQLFDRDMKKIEAFVGSKTIVQIRSHAQKYLFLKVOAKNCTREHVPDPKPKRASHPTPQKA 136
    |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
Db 76 LHLFDRDMKKIEAFVGSKTIVQIRSHAQKYLFLKVOAKNCTREHLPDPKPKRANHPTPQKA 135

QY 137 SK 138
    |
Db 136 PK 137

```

RESULT	9		
0943P3			
ID	0943P3;	PRELIMINARY;	PRT; 255 AA.
AC	0943P3;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	P001G08.21		
GN	P001G08.21.		
OS	Oryza sativa (Rice).		
OS	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta		
OC	Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae;		
OC	Ehrhartoideae: Oryzaceae: Oryza.		
OX	NCBI_TaxID=4530;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. NIPPONBARE;		
RA	Sasaki T., Matsumoto T., Yamamoto K.;		
RT	"Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC		
RT	clone:P001G08."		
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.		
DB	EMBL: A0003225; BABS649.1; -		
DB	SEQUENCE 255 AA; 28469 MW; B1B946F8E342A67E CRC64;		

[illegible]





DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)  
 DT 01-MAY-2000 (TReMBLrel. 13, last annotation update)  
 DE MYB-RELATED PROTEIN.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. IR-BB21;  
 RA Han F., Kilian A., Chen J.P., Kudrna D., Steffenson B., Yamamoto K.,  
 RA Matsumoto T., Sasaki T., Kleinjofcs A.;  
 RT \*Sequence analysis of a rice BAC covering the syntenous barley Rpg1  
 RT region.\*;  
 RL Genome 0:0-0(1999).  
 DR EMBL: AF149806; AAF00137.1; -.  
 SQ SEQUENCE 103 AA; 11895 MW; 96535D5F069B32EA CRC64;

Query Match 11.5%; Score 171.5; DB 10; Length 103;  
 Best Local Similarity 37.9%; Pred. No. 7.4e-08;  
 Matches 39; Conservative 12; Mismatches 19; Indels 33; Gaps 3;  
 OY 23 LVGMALPGGLAPNLSASVSASASSEDsAKKIRKPYTITKRSRESWSEQEHDK----- 72  
 DB 8 LIDSSSPGPGP-----GPEVEDGGGRYRKPKPYTITKRSRESWTDPPGGOVPMWRCSCKH 60  
 OY 73 ----FLEALQL-----FDRDMKKIEAFVGSKTYIQI 99  
 DB 61 PCAFFMWVPSYLFFAFFTCFYNNRDRDMKKIEAFVGSKTYIQY 103

Search completed: August 12, 2002, 10:11:40  
 Job time: 223 sec

1.

2.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:08:47 ; Search time 11.64 Seconds  
(without alignments)  
938.051 Million cell updates/sec

Title: US-09-640-211A-1076  
Perfect score: 1492  
Sequence: 1 MPMLAETRYRDSFETSGSS.....EWIQLILKLCCTHDEPHNL 282

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	8.8	2472	1 NCR2_MOUSE	Q9W422 mus musculu
2	119	8.0	2517	1 NCR2_MOUSE	Q9W422 mus musculu
3	116.5	7.8	1733	1 NCR2_MOUSE	Q9W422 mus musculu
4	105	7.0	1487	1 BLM_DROME	Q9W422 mus musculu
5	104.5	7.0	919	1 SYNP_HUMAN	Q9W422 mus musculu
6	102.5	6.9	562	1 TBR1_YEAST	Q9W422 mus musculu
7	101	6.8	692	1 EOMD_XENLA	Q9W422 mus musculu
8	99.5	6.7	705	1 FXP1_MOUSE	Q9W422 mus musculu
9	98.5	6.6	939	1 ST20_YEAST	Q9W422 mus musculu
10	98.5	6.6	942	1 ZDS2_YEAST	Q9W422 mus musculu
11	97.5	6.5	1593	1 AT12_HUMAN	Q9W422 mus musculu
12	96.5	6.5	444	1 FXP2_HUMAN	Q9W422 mus musculu
13	94.5	6.3	977	1 DLP3_RAT	Q9W422 mus musculu
14	94	6.3	437	1 PAX6_ORYLA	Q9W422 mus musculu
15	93	6.2	338	1 OTX3_BRARE	Q9W422 mus musculu
16	93	6.2	707	1 SPO_HUMAN	Q9W422 mus musculu
17	93	6.2	1273	1 WBL1_YEAST	Q9W422 mus musculu
18	93	6.2	1745	1 ZOI_MOUSE	Q9W422 mus musculu
19	92.5	6.2	437	1 PAX6_ORYLA	Q9W422 mus musculu
20	92.5	6.2	568	1 AP9_HUMAN	Q9W422 mus musculu
21	92.5	6.2	640	1 MYB_BOVIN	Q9W422 mus musculu
22	92.5	6.2	825	1 WDR1_YEAST	Q9W422 mus musculu
23	92.5	6.2	859	1 YDRB_SCHPO	Q9W422 mus musculu
24	92.5	6.2	872	1 FPL_MYTCO	Q9W422 mus musculu
25	92	6.2	677	1 FXP1_HUMAN	Q9W422 mus musculu
26	92	6.2	2363	1 SPCO_MOUSE	Q9W422 mus musculu
27	92	6.2	2453	1 NCR1_MOUSE	Q9W422 mus musculu
28	91.5	6.1	736	1 DVL2_HUMAN	Q9W422 mus musculu
29	91.5	6.1	1005	1 P532_HUMAN	Q9W422 mus musculu
30	91.5	6.1	1493	1 M3K1_RAT	Q9W422 mus musculu
31	91	6.1	668	1 YAF6_SCHPO	Q9W422 mus musculu
32	91	6.1	708	1 GBF_DICDI	Q9W422 mus musculu
33	91	6.1	940	1 MA24_SCHCO	Q9W422 mus musculu

34	91	6.1	1210	1 PER3_HUMAN	P56645 homo sapien
35	90.5	6.1	410	1 RID_DROME	Q24106 drosophila
36	90.5	6.1	600	1 RID_MOUSE	Q9W422 mus musculu
37	90.5	6.1	1505	1 CTT2_HUMAN	Q14529 homo sapien
38	90.5	6.1	3924	1 ANK2_HUMAN	Q01464 homo sapien
39	90	6.0	285	1 COAT_BPVV4	Q04755 lactococcus
40	90	6.0	951	1 SFR8_HUMAN	Q12872 homo sapien
41	90	6.0	1185	1 DRPL_HUMAN	P54259 homo sapien
42	89.5	6.0	531	1 SRS2_CANTR	Q12600 candida tro
43	89.5	6.0	551	1 GZFG_YEAST	P42944 saccharomyc
44	89.5	6.0	680	1 NNSP_RABIT	P27123 oryctolagus
45	89	6.0	527	1 ABP2_SCHPO	Q9W422 schizosach

## ALIGNMENTS

RESULT	ID	NCR2_MOUSE	STANDARD:	PRT:	2472 AA.
AC	Q9W422	Q9W422	Q9W422		
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Nuclear receptor co-repressor 2 (N-COR2) (silencing mediator of retinoic acid and thyroid hormone receptor) (SMRT) (Thyroid-, retinoic-acid-receptor-associated co-repressor) (TR3 receptor-associating factor) (TRAC).				
GN	N-COR2 OR SMRT.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid-10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).				
RC	TISSUE-Spleen, and Brain;				
RX	MEDLINE-99178941; PubMed-10077563;				
RA	Ordenlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;				
RT	"unique forms of human and mouse nuclear receptor corepressor SMRT.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA).				
RC	TISSUE-Embryo;				
RX	MEDLINE-9919215; PubMed-10097068;				
RA	Part E.J., Schiroen D.J., Yang M., Li H., Li L., Chen J.D.;				
RT	"SMRT, a silencing mediator for retinoid and thyroid hormone receptors, extends isoform that is more related to the nuclear receptor corepressor.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).				
CC	- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.				
CC	- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TRFB.				
CC	- SUBCELLULAR LOCATION: Nuclear.				
CC	- ALTERNATIVE PRODUCTS: 2 ISOFORMS, ALPHA (SHOWN HERE) AND BETA; ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY EMBRYOS.				
CC	- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2 AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION DOMAINS (ID1 AND ID2).				
CC	- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RAR. SEQUENCES FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR SPECIFICITY.				
CC	- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.				
CC	- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.				

CC -1- SIMILARITY: CONTAINS 2 CORN BOX.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF113001; AAD20944.1; -  
 DR EMBL: AF113002; AAD20945.1; -  
 DR EMBL: AF125671; AAD22972.1; -  
 DR MGI:1337080; Ncor2.  
 DR InterPro: IP001005; MYB\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 2.  
 DR SMART: SM00395; SANT; 2.  
 DR PROSITE: PS0090; MYB\_3; 1.  
 KM Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 KM Coiled coil; Alternative splicing  
 FT DOMAIN 155 207 COILED COIL (POTENTIAL).  
 FT 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).  
 FT DNA\_BIND 429 474 SANT-A (POTENTIAL).  
 FT MYB.  
 FT COILED COIL (POTENTIAL).  
 FT COILED COIL (POTENTIAL).  
 FT PRO-RICH.  
 FT PRO-RICH.  
 FT PRO-RICH.  
 FT CORN BOX OF 101.  
 FT CORN BOX OF 102.  
 FT POLY-GLN.  
 FT POLY-ALA.  
 FT MISSING (IN ISOFORM BETA).  
 FT M -> RL (IN REF. 2).  
 FT PPMALYDA -> RHVVR (IN REF. 2).  
 FT D -> H (IN REF. 1; AAD20944).  
 FT T -> A (IN REF. 1; AAD20944).  
 FT V -> A (IN REF. 2).  
 FT HHLPHRLMTRMNKKPRLLQLPRQMPRSRLRPRRSWE  
 FT -> PSPAAPATVVDKDECAAPAPAPQTEDEAKSEAEI  
 FT DVG (IN REF. 2).  
 FT E -> K (IN REF. 1; AAD20945).  
 FT E -> K (IN REF. 1; AAD20945).  
 FT E -> K (IN REF. 1; AAD20945).  
 FT S -> F (IN REF. 1; AAD20945).  
 FT I -> IO (IN REF. 1; AAD20944).  
 FT PKLPPEPMSGLPPI -> QSYRLSPHACHRLPSH  
 FT (IN REF. 2).  
 FT PHADPSA -> TRADPL (IN REF. 2).  
 FT MISSING (IN REF. 2).  
 FT MISSING (IN REF. 2).  
 FT G -> E (IN REF. 2).  
 FT GSATSGSTIKGUPSTGAAGPSTGSGITNG -> APPVEA  
 FT SPRASQPCGRPOQLRLVHR (IN REF. 2).  
 FT A -> S (IN REF. 2).  
 FT MISSING (IN REF. 2).  
 FT A -> P (IN REF. 2).  
 FT A -> G (IN REF. 2).  
 FT A -> G (IN REF. 2).  
 FT N -> S (IN REF. 2).  
 FT N -> G (IN REF. 2).  
 FT TA -> AV (IN REF. 2).  
 FT LE -> SK (IN REF. 2).  
 FT T -> A (IN REF. 2).  
 FT 2472 AA; 270856 MW; 2A58FADF7B72858 CMC64;  
 SQ SEQUENCE

Query Match

8.8%; Score 131; DB 1; Length 2472;

Best Local Similarity 23.0%; Pred. No. 0.15;  
 Matches 65; Conservative 34; Mismatches 141; Indels 42; Gaps 10;  
 QY 6 ETVRDSFETTCGSSVDLVGALPGLAPNLSASVSASASDASKKTRKRYTTIKRES- 64  
 Db 540 ELSEKEDTDDTSGEDNDEKAEVASKGRRTANSQGRKRITRSMANENHEETATPOOSSE 599  
 QY 65 -----WSFOEDKPLEALQLFDROWKTEAEVGSKTYIQLRSHAKYFLKQKN 113  
 Db 600 LASMENSSNWTBEMETAKKGLLEHGRNNSAIAKRVGSKTV---SCKNFYFNKKR 655  
 QY 114 GTREHVPPRRKRKASHIPYPOKASKNVPVSQVSTAPPTATQTD---SGYPR---AE 166  
 Db 656 QNLDIELLOHKLKMEKERNARRKKTPAAASETAPPAEDEMDSGASANEELAE 715  
 QY 167 SSILITKSSSCPTV-----SSWVHTIPSIDAFVEKDDGPGIETGNMCSG 216  
 Db 716 EAEASQASGNEVPVNGECSGPAVANNSSDITSVPS-PRSEATKRTGTGKPP--TGTALPA 771  
 QY 217 STESSPTWPPCSEITPEKVKPDEFQVYKFFIGSVFDPSTTDIL 258  
 Db 772 ATQ--PPVPPP--EPPAVAPAEPPVPDASGP-SPSPSHL 808  
 RESULT 2  
 NCR2\_HUMAN STANDARD: PRT: 2517 AA.  
 AC Q9Y618; Q9Y500; Q1354; Q00613; Q15416;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Nuclear receptor co-repressor 2 (N-COR2) (Silencing mediator of  
 DE retinoid acid and thyroid hormone receptor) (SMRT) (Thyroid-  
 DE retinoid-acid-receptor-associated co-repressor) (T3 receptor-  
 DE associating factor) (TRAC) (CTG26).  
 GN NCR2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).  
 RC TISSUE-Pituitary.  
 RA MEDLINE-9917841; PubMed-10077563;  
 RX Ordeulich P., Downs M., Xie W., Genin A., Splinter N.B., Evans R.M.;  
 RL "Unique forms of human and mouse nuclear receptor corepressor SMRT.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).  
 RN [12]  
 RN SEQUENCE FROM N.A. (ISOFORM SMRT).  
 RC TISSUE-cervical adenocarcinoma;  
 RA MEDLINE-99199215; PubMed-10097068;  
 RX Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;  
 RT "SMRte, a silencing mediator for retinoid and thyroid hormone  
 RT receptors, an extended isoform that is more related to the nuclear  
 RT receptor corepressor.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).  
 RN [13]  
 RN SEQUENCE OF 1023-2517 FROM N.A.  
 RC TISSUE-cervical adenocarcinoma;  
 RA MEDLINE-96008552; PubMed-756127;  
 RX Chen J.D., Evans R.M.;  
 RT "A transcriptional co-repressor that interacts with nuclear hormone  
 RT receptors.";  
 RT Nature 377:454-457(1995).  
 RN [14]  
 RN SEQUENCE FROM N.A. (ISOFORM TRAC-1).  
 RC TISSUE-Petal liver;  
 RA MEDLINE-96408715; PubMed-8813722;  
 RX Sande S., Privalsky M.L.;  
 RT "Identification of TRACs (T3 receptor-associating cofactors), a family  
 RT of cofactors that associate with, and modulate the activity of,  
 RT nuclear hormone receptors.";  
 RT Mol. Endocrinol. 10:813-825(1996).



RN [5]  
 RP SEQUENCE OF 428-613 FROM N.A.  
 RC TISSUE-Brain cortex:  
 RX MEDLINE:97369492; PubMed:9225980;  
 RA Margolis R.L., Adoriam M.R., Gatchell S.B., Li S.H., Kidwai A.S.,  
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
 RT "CDNAs with long CAG trinucleotide repeats from human brain.";  
 RL Hum. Genet. 100:114-122(1997).  
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIB.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: SMRT/TRAC-2 (SHOWN HERE) AND  
 CC TRAC-1; ARE PRODUCED BY ALTERNATIVE SPLICING. TRAC-1 CONTAINS ONLY  
 CC THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN  
 CC ANTI-REPRESSOR.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS. HIGH LEVELS OF EXPRESSION ARE  
 CC DETECTED IN LUNG, SPLEEN AND BRAIN.  
 CC -1- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.  
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES  
 CC FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 MIB-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CORN BOX.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC  
 DR EMBL, AF113003; AAD20946.1; -;  
 DR EMBL, AF125672; AAD22973.1; -;  
 DR EMBL, U37146; AAC50236.1; -;  
 DR EMBL, S83390; AAB50847.1; -;  
 DR EMBL, U80750; AAB91446.1; -;  
 DR MIM: 600848; -;  
 DR InterPro: IPR001005; MYB\_DNA\_bind.  
 DR Pfam: PF00249; myb\_DNA-binding; 2.  
 DR SMART: SM00395; SANT; 2.  
 DR PROSITE: PSS0090; MYB\_3; 1.  
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;  
 KW Coiled coil; Alternative splicing.  
 FT DOMAIN 174 215  
 FT 254 312 COILED COIL (POTENTIAL).  
 FT 439 474 INTERACTION WITH SIN3A/B (BY SIMILARITY).  
 FT 613 657 SANT-A (POTENTIAL).  
 FT 778 820 MYB.  
 FT 820 820 COILED COIL (POTENTIAL).  
 FT 2139 2143 PRO-RICH.  
 FT 2346 2346 CORN BOX OF ID1.  
 FT 494 510 CORN BOX OF ID2.  
 FT 510 510 POLY-GLN.  
 FT 685 685 POLY-LYS.  
 FT 994 1002 POLY-PRO.  
 FT 1384 1389 POLY-PRO.  
 FT 1842 1846 POLY-GLY.  
 FT 2479 2482 POLY-PRO.  
 FT DOMAIN

FT VARSPIC 1 1702 MISSING (IN ISOFORM TRAC-1).  
 FT VARSPIC 2353 2398 MISSING (IN ISOFORM TRAC-1).  
 FT 295 295 L -> P (IN REF. 2).  
 FT 295 295 K -> E (IN REF. 2).  
 FT 309 309 L -> W (IN REF. 2).  
 FT 309 309 MISSING (IN REF. 2).  
 FT 352 352 A -> P (IN REF. 2).  
 FT 365 365 SS -> EF (IN REF. 5).  
 FT 612 612 S -> T (IN REF. 2).  
 FT 711 711 MISSING (IN REF. 2).  
 FT 724 740 MISSING (IN REF. 2).  
 FT 740 796 RTSRAPIEP -> PEDIPAPTES (IN REF. 2).  
 FT 804 804 G -> L (IN REF. 2).  
 FT 814 814 S -> F (IN REF. 2).  
 FT 817 817 A -> S (IN REF. 2).  
 FT 889 889 G -> R (IN REF. 2).  
 FT 1023 1030 SRSPAPPA -> MEAMPAHP (IN REF. 3).  
 FT 1034 1034 A -> AKRPVEFA (IN REF. 2).  
 FT 1894 1894 K -> T (IN REF. 4).  
 FT 2494 2494 P -> A (IN REF. 4).  
 SQ SEQUENCE 2517 AA; 274031 MW; F5805C01761258C0 CRC64;  
  
 Query Match 8.0%; Score 119; DB 1; Length 2517;  
 Best local Similarity 20.2%; Pred. No. 1.1;  
 Matches 57; Conservative 33; Mismatches 134; Indels 58; Gaps 8;  
  
 QY 9 RDSFTTSGGSSVDLVGALPLGIALPNLSASVSASDSAKKIRKPYTTIKSNE----- 64  
 DB 547 KEKDTDTSGEDNDEKEAASKGRKRTANSQGRKGRITFSMANENANSEERATIPQOASAEIAS 606  
 QY 65 -----WSBOEHKFLFEALOLFPODMKIEAFVGSVTYQIISHAKYFLKVKQKNGTR 116  
 DB 607 MELNASSWTEDEEMETAKKGLLEHGRNMSAIAKMGSTV-----SQCNKFTFNKKKRN 662  
 QY 117 EHVPPRPKRKASHPEYPOKASKNPVSOOVSTAFPTA-----ATOLDSCGYPRAESSS 169  
 DB 663 DEILQOHKIMKEKRNARRKKKAPAASEAFPPVEDEMEASGVSGNEBEKVEAE 722  
 QY 170 ILTKSGSCP-----TV-SSWVHTITPSIDASFVKD-----DGGRPG 206  
 DB 733 ALHAGSNEVPRGECSPATVNNSSDTEISIPNHE-AKKDVGONGKPPRATGADGCPGG 781  
 QY 207 IET-----GNNCSGSTESSPTWPCSEIRPKVMP 237  
 DB 782 PPTPRKRSRAPIETPPASSEATGAPTPPPAPPSAPPPVVP 823  
  
 RESULT 3  
 VNUA\_PRIVKA STANDARD; PRT: 1733 AA.  
 ID VNUA\_PRIVKA  
 AC P334B5;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Probable nuclear antigen.  
 OS Pseudorabies virus (strain Kaplan) (PRV).  
 OC Viruses; dsDNA viruses, not RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirus.  
 OX NCBI\_TaxID=33703;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021039; PubMed=2171211;  
 RA Vleck C., Kozmik Z., Paces V., Schlim S., Schwyzer M.;  
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
 RT oriented open reading frame: characterization of their promoter and  
 RT enhancer regions.";  
 RL Virology 179:365-377(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M34651; AAA47471.1; -

DR PIR; B45344; B45344. -

FT DOMAIN 112 117 POLY-THR.

FT DOMAIN 179 1733 GLY-RICH.

FT DOMAIN 192 196 POLY-SER.

FT DOMAIN 271 298 POLY-PRO.

FT DOMAIN 304 308 POLY-ARG.

FT DOMAIN 883 889 POLY-GLY.

FT DOMAIN 1398 1405 POLY-GLY.

SO SEQUENCE 1733 AA; 172166 MW; 0C8CB8BE475B5E2 CRC64;

Query Match 7.8%; Score 116.5; DB 1; Length 1733;  
 Best Local Similarity 26.8%; Pred. No. 1.1;  
 Matches 38; Conservative 15; Mismatches 56; Indels 33; Gaps 6;

QY 116 REHVPRPRPKKASH-PYQKASKNV--PVSOQVSTAFPTAATQLOSGYV-----PRA 165  
 DB 93 RDHLPRRTRRDQHRRPPTTTTITIKDPQHPDPLLPKTLQEDDPHLRPTRPDPSA 152

QY 166 ESSSILTKSSGSCPTVSSWVHTIPSDASFVEKKDQSGPGETGNMNCSSGSESSPPT- 224  
 DB 153 KTHHHQDPGCGPSTSSHHHNDP-----PGGGPPSPRRPSTSSSHSQGPST 203

QY 225 -----WPQSEIPEKV 235  
 DB 204 RPPPPQRRPPRRWPPPS--PQKI 223

RESULT 4  
 BLM\_DROME STANDARD; PRT; 1487 AA.

AC 09VG18; 09Y062;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE Bloom's syndrome protein homolog (EC 3.6.1.-) (Dmblm) (Mutagen-  
 sensitive protein 309) (RecQ helicase homolog).  
 GN mus309 or blm or CG6920.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OX NCBI\_TaxID=7227;  
 RN RP SEQUENCE FROM N.A., AND REPEATS.  
 RC STRAIN=Canton-S;  
 RX MEDLINE=99160561; PubMed=10049920;  
 RA Kusano K., Berres M.E., Engels W.R.;  
 RT "Evolution of the RECQ family of helicases: a Drosophila homolog,  
 RT Dmblm, is similar to the human Bloom syndrome gene.";  
 RL Genetics 151:1027-1039(1999).  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Beus P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,  
 RA Burks D.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferltera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeagwa C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spiter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sviders R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: Participates in DNA replication and may participate in  
 CC repair. Exhibits a magnesium dependent ATP-dependent DNA-helicase  
 CC activity that unwinds single- and double-stranded DNA in a 3'-5'  
 CC direction (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RECQ SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HRDC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use, by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U92536; A041441.1; -  
 DR EMBL; AF003692; AAF54691.1; -  
 DR Flybase; FBgn002306; mus309.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002464; DEAH\_ATP\_helicase.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR002121; HRDC.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00570; HRDC; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 DR SMART; SM00341; HRDC; 1.  
 DR PROSITE; PS006590; DEAH\_ATP\_HELICASE; 1.  
 KW Hydrolyase; Helicase; ATP-binding; DNA replication;  
 KW Repeat; Nuclear Protein.  
 FT DOMAIN 89 138  
 FT  
 FT REPEAT 89 112  
 FT REPEAT 115 138  
 FT DOMAIN 225 230  
 FT DOMAIN 464 467  
 FT DOMAIN 1283 1363  
 FT DOMAIN 1369 1372  
 FT DOMAIN 1416 1432  
 FT NP\_BIND 759 766  
 FT SITE 865 868  
 FT CONFLICT 98 98  
 FT CONFLICT 110 110  
 FT CONFLICT 126 126  
 FT CONFLICT 134 136  
 FT  
 FT 2 x 24AA REPEATS OF L-D-L-S-V-S-P-L-A-E-  
 FT L-[SP]-A-K-K-K-[YS]-[AD]-R-D-[SP]-P-P-K-  
 FT P.  
 FT 1.  
 FT 2. POLY-PRO.  
 FT POLY-SER.  
 FT HRDC.  
 FT POLY-GLU.  
 FT NUCLEAR LOCALIZATION SIGNAL (BY  
 FT SIMILARITY).  
 FT ATP (BY SIMILARITY).  
 FT DEAH BOX.  
 FT K -> E (IN REF. 1).  
 FT L -> P (IN REF. 1).  
 FT L -> P (IN REF. 1).  
 FT SPK -> PPP (IN REF. 1).

FT	CONFLICT	169	169	O -> P (IN REF. 1).
FT	CONFLICT	222	222	P -> S (IN REF. 1).
FT	CONFLICT	299	299	Y -> N (IN REF. 1).
FT	CONFLICT	417	417	M -> V (IN REF. 1).
FT	CONFLICT	459	459	Q -> R (IN REF. 1).
FT	CONFLICT	471	471	S -> C (IN REF. 1).
FT	CONFLICT	496	496	H -> P (IN REF. 1).
FT	CONFLICT	594	594	D -> G (IN REF. 1).
FT	CONFLICT	614	614	G -> A (IN REF. 1).
FT	CONFLICT	805	805	K -> E (IN REF. 1).
FT	CONFLICT	1116	1116	E -> G (IN REF. 1).
SO	SEQUENCE	1487	AA: 166077	MM: 0761B8005E29432 CRC64:

[illegible]

DR	SMART: PF00023; ank: 4.
DR	SMART: SM00248; ANK: 2.
DR	PROSITE: PS50088; ANK_REPEAT: 1.
DR	PROSITE: PS50297; ANK_REPEAT_REGION: 1.
KW	Coiled coil; Repeat; ANK repeat.
FT	REPEAT 349 380 ANK 1.
FT	REPEAT 384 413 ANK 2.
FT	REPEAT 419 448 ANK 3 (POTENTIAL).
FT	REPEAT 456 485 ANK 4.
FT	DOMAIN 515 552 COILED COIL (POTENTIAL).
FT	REPEAT 603 632 ANK 5 (POTENTIAL).
FT	REPEAT 659 729 ANK 6 (POTENTIAL).
SD	SEQUENCE 919 AA: 100380 MW: 350022250024E8F4 CRC64:

```

Query Match Similarity      7.0%; Score 104.5; DB 1; Length 919;
Best Local Similarity       24.0%; Pred. No. 3.7;
Matches    69; Conservative   31; Mismatches   110; Indels    77; Gaps     12;

OY      22 DLVGNALPGLAPN--LSSASVSASA-----SEDSAKIRKPYTTTKRESNEDEHDKF 73
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      582 DSVANSRKGVEGDIQVLGLSLSSASSRRARRKANEDSDKLRLQLRGKEISENVCTOKLSLE 641
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      74 LEALDLPDMKKIDIAFGSKTVIOIRSHAQVFYLKYONKNGREHVPPRPKRKSHYP 133
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      642 FQDAQAASSRNRSKRIILEKRELKLARLQMQRSLSESOTDSNNSDPRPTTPYRKARDRRP 701
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      134 OKASKRVNPVSQQOVSTAFPAATATOLDGGYYPRAESSSILTRSGSGSCPVTSSVMYNHTIP--- 190
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      702 Q-----DIVESVES-----MDS-----AESIHLMIKK-----HTLASGG 730
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      191 -----SIDASFEXKDGGPGPIETGNCCSGSTESS-----PRTWPCSEI 231
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      731 RRFPSIRAN--KSIDGHSP-----SPTSSESPDLSEQYTGSGSIFPNQP--SGD 777
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      232 PEKVAPDPESQVYKFTGSVFPDPTSDHLKLKLEMWLILKLCCTHEEP 278
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      778 PQQPFSPDSTAQKVATS---PKSAALKSPSSKRRRTSQNLKLRVTFEFP 821
           | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      6
TBFL_YEAST TBFL_YEAST STANDARD: PRT; 562 AA.
AC          002457;
DT          01-JUL-1993 (Rel. 26, created)
DT          01-OCT-1996 (Rel. 34, last sequence update)
DT          16-OCT-2001 (Rel. 40, last annotation update)
DE          TBFL protein (TTAGCG repeat-binding factor 1) (TBF alpha).
GN          TBF1 OR YPL128C OR LPI1BC.
OS          Saccharomyces cerevisiae (Baker's yeast).
OC          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC          Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX          NCBI_Taxid=4932;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=LN224;
RX          MEDLINE=93140769; PubMed=8423796;
RA          Brigati C., Kurtz S., Balderes D., Vidali G., Shore D.M.;
RT          "An essential yeast gene encoding a TTAGCG repeat-binding protein.";
RL          Mol. Cell. Biol. 13:1306-1314(1993).
RN          [2]
RP          SEQUENCE FROM N.A.
RA          Hall J., Depaulo T., Ahmed A., Bussey H., Fortin N., Frieseen J.D.,
RA          Storms R.K., Vo D.H., Wang Y., Winnett E.;
RL          Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
CC          -!- FUNCTION: BINDS TO TTAGCG REPEATS. ESSENTIAL FOR CELL GROWTH.
CC          MAY PLAY A ROLE IN TELOMERE FUNCTION AND/OR STRUCTURE. MAY
CC          REGULATE GENE EXPRESSION.
CC          -!- SUBCELLULAR LOCATION: Nuclear.
CC          -!- SIMILARITY: TO S.POMBE SPEC1967.13.
CC          -!- SIMILARITY: CONSTRAINTS 1 MYB-LIKE DOMAIN.
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X65394; CAA49191.1; -  
 DR EMBL: U43703; AAB68230.1; -  
 DR PIR: S28558; S28558.  
 DR PIR: A48079; A48079.  
 DR TRANSFAC: T01246; -  
 DR SGD: S0006049; TRF1.  
 DR Pfam: PF00249; myb\_DNA-binding; 1.  
 DR SMART: SM00395; SMT; 1.  
 DR PROSITE: PS50090; MYB\_3; 1.  
 DR Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 404 456 MYB.  
 FT CONFLICT 71 71 E -> R (IN REF. 1).  
 FT CONFLICT 371 373 DAA -> ERR (IN REF. 1).  
 FT SEQUENCE 562 AA; 62823 MW; 2340F086468EC54F CRC64;

Query Match 6.9%; Score 102.5; DB 1; Length 562;

Best Local Similarity 20.7%; Pred. No. 2.9; Indels 55; Gaps 7;  
 Matches 45; Conservative 38; Mismatches 79;

OY 29 PGLAPNLSASVSASASEDSAKK-----IRPYRTTKRSNSEQEHD 71  
 DB 363 PALQSIYDAVAASMSNSSGPHSHNSNSNNNSIDLRKPK---KAKPRMSKREEE 418  
 OY 72 KFLFALQLFDRDKKIEAFV-----SKTVDIRSHAKYFLKVKNGTREHVP 121  
 DB 419 ALVEGKLEVPSPMSKIDLYGPGGKITENLKNRTQVCLKRKARMKQLYKSG-----K 472  
 OY 122 PRPKKASHPYPOKASNVPSQOVSTAPPTATOLDGYYPRAESSTILTKSGSCP-T 180  
 DB 473 PLP-----DYLKVTGNLEKTYAKKKFSQS-----PNSSTIMEÖNLQHPSS 515  
 OY 181 VSSWVHHTIPSIDASPEVKDDGPPGIGTGNCSGS 217  
 DB 516 AASATEDTQTHQEDSHGONSDNMPNSNGLFGNSTSDNT 552

RESULT 7  
 EOMD\_XENLA STANDARD; PRT; 692 AA.  
 AC P79944;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Eomesodermin.  
 GN EOMES.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;  
 OC Anopodidae; Xenopus.  
 OC NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Gastrula;  
 RX MEDLINE=97133207; PubMed=8978604;  
 RA Ryan K., Garrett N., Mitchell A., Gurdon J.B.;  
 RT "Eomesodermin, a key early gene in Xenopus mesoderm differentiation.";  
 RL Cell 87:989-1000(1996).  
 CC -1- FUNCTION: INVOLVED IN MESODERM DIFFERENTIATION. ACTIVATES WNT8,  
 CC BRACHYURY, CHD AND MIX.1 EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED AT OR JUST AFTER MIDLASTULA  
 CC TRANSITION (STAGE 8). MAXIMALLY EXPRESSED AT STAGE 10 AS AN  
 CC EQUATORIAL MESODERM BAND, MORE PROMINENTLY ON THE DORSAL SIDE  
 CC AND AROUND THE INVAGINATING DORSAL LIP.  
 CC -1- INDUCTION: BY ACTIVIN.

CC -1- DOMAIN: CONTAINS 13 S-P-X-X REPEATS.  
 CC -1- SIMILARITY: CONTAINS 1 T-BOX DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U75996; AAC60061.1; -  
 DR HSSP: P24781; 1XBR.  
 DR InterPro: IPR001699; T-box.  
 DR Pfam: PF00907; T-box; 1.  
 DR PRINTS: PR00937; TBOX.  
 DR SMART: SM00425; TBOX; 1.  
 DR PROSITE: PS01283; TBOX\_1; 1.  
 DR PROSITE: PS01264; TBOX\_2; 1.  
 DR PROSITE: PS50252; TBOX\_3; 1.  
 DR Developmental protein; Transcription regulation; DNA-binding;  
 KW Nuclear protein; Repeat.  
 FT DNA\_BIND 263 443 T-BOX.  
 FT SEQUENCE 692 AA; 75943 MW; 9D129A67F6357989 CRC64;

Query Match 6.8%; Score 101; DB 1; Length 692;

Best Local Similarity 21.4%; Pred. No. 4.7;  
 Matches 63; Conservative 37; Mismatches 95; Indels 100; Gaps 13;

OY 5 AETVRDSFETTSGGSSVDYGNALPGLAPNLSASVSASASEDSAKKIKPPTTKSRPS 64  
 DB 438 AKGFRNDYD-----SMYTASESDRLTPSPADSPSHOIVGTRRS 477  
 OY 65 MSEQEHKFLFALQLFDRDKKIEAFVSGKTYIQLSHAKYFLKVKNGTREHVP 123  
 DB 478 VQFFPDQOVNNL-----PRARYSGRTVPQANG-----LISQTEEVANPPQKW 525  
 OY 124 ---PKRKA-----SHP---YQKRA-----SK 138  
 DB 526 FVTPVQQAAMANKLDMGAYETDYSGLLTYYGKSLPIQSHPMAYYPPDAAFASMGWGSR 585  
 OY 139 NVPSQOVSTAPPTATOLDGYYPRAESSTILTKSGSCPVSVMVHTTISDASPE 198  
 DB 586 GSTRYQRRKMTTSLPWSRSSPSCF-----SEDLPLPKDKVKEEMSSWV-ETPSTISL--- 636  
 OY 199 KDDGCGPIETGN-----NCSGSTRSSPTMPPCSEIPEKVKPDFSOVYKFG 247  
 DB 637 --DSNDSGYTTACKRRRLSPSTSSNENSPPI--KCEDICTE---DYKDATKGLG 684

RESULT 8  
 FXP1\_MOUSE STANDARD; PRT; 705 AA.  
 AC P58462;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Forkhead box protein P1 (Forkhead-related transcription factor 1).  
 GN FOXP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (TSOFORMS A, B AND C).  
 RC STRAIN=C57BL/6; TISSUE=Lung;  
 RX MEDLINE=21347947; PubMed=11358962;  
 RA Shu W., Yang H., Zhang L., Lu M.M., Morrisey E.E.;  
 RT "Characterization of a new subfamily of winged-helix/forkhead (Fox)  
 RT genes that are expressed in the lung and act as transcriptional  
 RT repressors."; J. Biol. Chem. 276:27488-27497(2001).

```

CC -1- FUNCTION: Transcriptional repressor that play an important role in
CC the specification and differentiation of lung epithelium.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; A (shown here), B and C, are
CC produced by alternative splicing. The isoform C is produced by the
CC use of alternative initiation codons in the same reading frame.
CC -1- TISSUE SPECIFICITY: Highest expression in the lung, brain, and
CC spleen. Lower expression in heart, skeletal muscle, kidney, small
CC intestine (isoform C not present) and liver.
CC -1- DEVELOPMENTAL STAGE: Expressed in developing lung, neural,
CC intestinal and cardiovascular tissues.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AF393103; AAK69648.1; -
CC DR EMBL; AF393104; AAK69649.1; -
CC DR EMBL; AF393105; AAK69650.1; -
CC DR MGD; MGI:191404; Foxp1.
CC DR PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
CC DR PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
CC DR PROSITE; PS00659; FORK_HEAD_3; 1.
CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
CC DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; FALSE_NEG.
CC KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
CC Nucleic protein; Alternative splicing; Alternative initiation.
CC FT DNA_BIND 493 583
CC FT ZN_FING 334 359
CC FT DOMAIN 55 60
CC FT DOMAIN 71 107
CC FT DOMAIN 161 164
CC FT VARSPIC 1 250
CC FT VARSPIC 539 602
CC FT VARSPIC 539 602
CC FT SEQUENCE 705 AA; 78833 MW; 92962B82917C079D CMC64;
CC -----
Query Match 6.7%; Score 99.5; DB 1; Length 705;
Best local Similarity 22.8%; Pred. No. 6.2;
Matches 50; Conservative 26; Mismatches 92; Indels 51; Gaps 7;
OY 13 EFTSGG-SSVDLVGALDELGLAPNLSSASVSASASEDSAKKIKPTTTSRESKWSBQEND 71
Dy 271 ETTSSNHSSLDLTSTCVSSASAPKSSLLMNPASTNGQLSVTP-----KRSLSHEHP 325
OY 72 -----KFLMALQFLFRDMRKLEAFVSGSKTVIQTIRSHQ 104
Dy 326 HSHPLTGHGVCCKWPCGEAVCDDFPAFLKHLNHEHALD-DRSTAQCRVQWQVQQLFLQLA 384
OY 105 KYFLKVGQKNGREHVPPRPKRASHPPYQKASKVNPVSQGVSTA---FPTAAQLQDSG 160
Dy 385 KDKERLQAMMTLHVKSTEPK---AAPQGLNLVSSVTSLSKASASAPSLPHTPTTPAP 441
OY 161 YTPRAESSILTKSGSSSCTPVSSWVHHTIPSIDASFVEK 199
Dy 442 LTPVTQGPSVITTTSM-----HHVGPTRRRYS DK 470
-----
RESULT 9
ST20_YEAST STANDARD: PRT: 939 AA.
AC Q003497:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase STE20 (EC 2.7.1.-).
GN STE20 OR YHL007C.

```

OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93099855; PubMed=1464311;  
 RA Leberer E., Dignard D., Hartsus D., Thomas D.Y., Whiteway M.;  
 RT "The protein kinase homologue Ste20p is required to link the yeast  
 RT pheromone response G-protein beta gamma subunits to downstream  
 RT signalling components.";  
 RL EMBL J. 11:4815-4824(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93133807; PubMed=8421676;  
 RA Ramer S.W., Davis R.W.;  
 RT "A dominant truncation allele identifies a gene, STE20, that encodes  
 RT a putative protein kinase necessary for mating in Saccharomyces  
 RT cerevisiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:452-456(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Kuc Z., Faveil A., Fulton L., Johnson S., Geisel C., Kirsten J.,  
 RA Dueba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,  
 RA Vaudin M.;  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT VIII.";  
 RL Science 265:2077-2082(1994).  
 CC -I- FUNCTION: REQUIRED TO LINK THE PHEROMONE RESPONSE G-PROTEIN BETA  
 CC GAMMA SUBUNITS TO DOWNSTREAM SIGNALING COMPONENTS. IT IS THOUGHT  
 CC THAT IT CAN PHOSPHORYLATE STE5, NEEDED FOR MATING IN HAPLOID  
 CC CELLS, INDUCTION OF A MATING-SPECIFIC GENE FUS1, INDUCTION OF  
 CC MATING-SPECIFIC MORPHOLOGIES, AND PHEROMONE-INDUCED PROLIFERATION  
 CC ARREST. PHOSPHORYLATES STE11.  
 CC -I- PATHWAY: RESPONSE TO PHEROMONE-INDUCED SIGNAL.  
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -I- PTM: AUTOPHOSPHORYLATED ON SERINE RESIDUES.  
 CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC STE20 SUBFAMILY.  
 CC -I- SIMILARITY: CONTAINS 1 GBD DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, M94719; AAA35111.1; -;  
 DR EMBL, L04655; AAA35038.1; -;  
 DR EMBL, L04655; AAA35039.1; -;  
 DR EMBL, U11581; AAB69747.1; -;  
 DR PIR, S28394; S28394.  
 DR PIR, S46821; S46821.  
 DR HSP, P11362; 1P6K.  
 DR SGD, S0000999; STE20.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR000095; PAK\_box\_P11\_Rho\_binding.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR Pfam: PF00786; PBD; 1.  
 DR SMART: SM00285; PBD; 1.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR PROSITE: PS0108; GBD; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE: PS00108; PROTEIN\_KINASE\_ST. 1.  
 KW Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 KM Phenomenon response; Phosphorylation.  
 FT DOMAIN 337 395  
 FT NP\_BIND 620 871 PROTEIN KINASE.  
 FT BINDING 626 634 ATP (BY SIMILARITY).  
 FT ACT\_SITE 739 739 BY SIMILARITY.  
 FT CONFLICT 19 19 N -> S (IN REF. 2).  
 FT CONFLICT 134 134 I -> M (IN REF. 2).  
 FT CONFLICT 271 271 P -> S (IN REF. 2).  
 SO SEQUENCE 939 AA; 102362 MW; 69C1C12F5B87733C CRC64;

Query Match 6.6%; Score 98.5; DB 1; Length 939;  
 Best Local Similarity 20.8%; Pred. No. 10;  
 Matches 69; Conservative 44; Mismatches 123; Indels 95; Gaps 13;

QY 10 DSFETSGSSVDLVGM-----ALPGLAPNLSSASVSASDSAKKIRKP 55  
 DB 90 DPDPTRVSSSVISGSSSSPHSNIDETKSLAVPNTNTITPDHSDN----- 142  
 QY 56 YTIKRSRESSEOEH---DKFLALOLF-----RDMKIEAFVGSKVYQIRS--- 101  
 DB 143 ---TFSTINASESDHOFMDTLKSLTLDSTFTENNATVYKHQCPVASSSTVNSKSTDI 199  
 QY 102 -----HAQYFLKVNKNGTREVHPPRPKRKASHPYQKASK 138  
 DB 200 RRAATPVSTPVYISKPSMTTTPROINSASHSLNPKH--KQHPKPKVPSKPEAKSPVSVKK 257  
 QY 139 NVPSVQGVSTAFPTAATOLDGCVYBRAESSILTKSGSCPTVSSWNHTIPSIDAFVE 198  
 DB 258 SFPKSNPLKNSNP-PKQOTEKSTY-----SSSKKRRSGSNSCTL-----RKMDVFTSFVQ 307  
 QY 199 -----KDDGGRPGIETGNCCSGSTSSPPTWPCSEIPEKVPKDFESQYKFTGSV-FD 251  
 DB 308 NIKRNSQDD-----KRASSSSNNSS-----SSSTALRLTSPYNNKHHIHGVGD 352  
 QY 252 PSTTDHLKLLKLEWLIQILKLCCTHEPPHNL 282  
 DB 353 SKTGEYTGLEPEWEKLLTSSGISRQEQQNM 383

RESULT 10  
 ZDS2\_YEAST  
 ID ZDS2\_YEAST STANDARD; PRT; 942 AA.  
 AC PS4786;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ZDS2 protein.  
 GN ZDS2 OR MCS1 OR YML109W OR YW8339.10.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96413276; PubMed=8816438;  
 RA Yu Y., Jiang Y.W., Wellinger R.J., Carlson K., Roberts J.M.,  
 RA Stillman D.J.;  
 RT "Mutations in the homologous ZDS1 and ZDS2 genes affect cell cycle  
 RT progression";  
 RL Mol. Cell. Biol. 16:5254-5263(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Skellton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases  
 CC -!- FUNCTION: NOT KNOWN; MULTICOPY SUPPRESSOR OF A VARIETY OF DEFECTS.  
 CC SEEMS TO INTERACT WITH CDC42. MAY PLAY AN IMPORTANT ROLE IN CELL  
 CC CYCLE PROGRESSION.  
 CC -!- MISCELLANEOUS: ZDS MEANS 'ZILLION DIFFERENT SCREENS' AS BOTH ZDS1

CC AND ZDS2 HAVE BEEN FOUND BY A WIDE VARIETY OF GENETIC SCREENS.  
 CC -!- SIMILARITY: STRONG, TO YEAST ZDS1/NC1/CE1.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 CC EMBL: U32938; AAB37541.1; -;  
 DR EMBL: 249210; CA89109.1; -;  
 DR SCD: S0004377; ZDS2.  
 FT CONFLICT 7 7 M -> V (IN REF. 1).  
 FT CONFLICT 23 23 T -> S (IN REF. 1).  
 FT CONFLICT 530 530 A -> P (IN REF. 1).  
 FT CONFLICT 546 546 A -> G (IN REF. 1).  
 FT CONFLICT 633 633 V -> A (IN REF. 1).  
 FT CONFLICT 668 668 N -> S (IN REF. 1).  
 FT CONFLICT 711 711 V -> VNC (IN REF. 1).  
 FT CONFLICT 723 723 MISSING (IN REF. 1).  
 FT CONFLICT 729 729 A -> E (IN REF. 1).  
 FT CONFLICT 833 833 S -> P (IN REF. 1).  
 FT CONFLICT 914 942 TGDIAFGDSALGMDKNDSDGTLILPDI -> HMRYSLQW  
 (IN REF. 1).  
 SO SEQUENCE 942 AA; 105495 MW; A1C9DD9A539E4291 CRC64;

Query Match 6.6%; Score 98.5; DB 1; Length 942;  
 Best Local Similarity 23.5%; Pred. No. 10;  
 Matches 73; Conservative 37; Mismatches 121; Indels 79; Gaps 15;

QY 6 EYR-----DSFETSGSSVDLVGMALPGLAPNLSSASVSASDSAKKIRKP-----R 53  
 DB 549 QTYQQLDDEDECEVYNEKADFPVNLSPRAKSTKRASBRINRKRSPIDQISEBA 608  
 QY 54 KPTITKRSRESSEOEHDK-----FLEALOLFPRDMKRTIAFGSKTVYQIRSHAO----- 104  
 DB 609 KSVYITPSPVSSSESQSPKFTAPAYVE-----KKVEL-----PTDQASTHKRNSLE 655  
 QY 105 -----KYFLKVNKNGT-----REHVPVPRPKRASH-----PYQKASKNVVSVQ 145  
 DB 656 KRLAKLFRKQHGNGTCKSDVYIKKSV-KKELKKKASHSLSKFRKSPKKKPEALEVERP 714  
 QY 146 VSTAEPPTAATOLD--SGYVRAESSILTKSGSCPTVSS--WVHHTIPSIDAFVEKDDG 202  
 DB 715 SSPTKTTTTPEDIPYASVTEPEVRSSNASTLLPDSHTSHSEPFVEFTISELDGDSFDISG 774  
 QY 203 GPGG--TEGNNGSSGT-----ESSPTWPCSEIPEK-----YKPDFSQ 241  
 DB 775 GDVNYDVEVHSSIRDTTLAGLEDIGAEKEDNTPSPAPQISTLPKRLTFEDVYKFDYSN 834  
 QY 242 V-YKFTGSVP 250  
 DB 835 APIKFTDSAF 844

RESULT 11  
 AT12\_HUMAN  
 ID AT12\_HUMAN STANDARD; PRT; 1593 AA.  
 AC P58397;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and  
 DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-  
 DE TS12).  
 GN ADAMTS12.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;

FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SO	SEQUENCE	1593 AA;	177545 MM;	07F9FA8B53BD83A3	CNC64;

  

Query Match	6.5%;	Score 97.5;	DB 1;	Length 1593;
Best Local Similarity	24.0%;	Pred. No. 22;		
Matches 53;	Conservative 27;	Mismatches 80;	Indels 61;	Gaps

  

QY	34	NLSASVSAASA-----	SEDSAKTRKRYTTTTSKRESSEDEHDKFLEALQFLDRDKKTE	88		
Db	951	NMSDCSVSCGGVIRVIRVETCAKKNDECDVTRKPN-----	FALCGLD-----	9933		
QY	89	AFVSGKTYIQRHMAQKXYFLKVKQ-----	NG-----	TREHVRPP-----	RPRKASHPYPOKAK	138
Db	994	-----	QCPSSRRVLKPNKGTISNGKNRPPTLKPVRPPSRPMYLTTPGPESMST	104		
QY	139	NVPSQGVSTAFPRATQ-----	LDSGYPRASSILITKSGSCPTVSSWHNTI	189		
Db	1043	STPA-----	ISSPPTAKEDLGKQKQODSSTOPELSSRLITSGTSQPIILTSLSLIG	109		
QY	190	PSIDASFEVKDGGPG-----	ITGNCCSGSTESSPTMP	226		
Db	1100	PSEEN--	VSSSDTGPTEGCGLVATTTSGSGLSSSRNPITMP	1138		

  

RESULT 12	FXF2_HUMAN	STANDARD;	PRT;	444 AA.
AC	Q12947; Q9U085;			
DT	01-NOV-1997 (Ref. 35, Created)			
DT	30-MAY-2000 (Ref. 39, Last sequence update)			
DT	16-OCT-2001 (Ref. 40, Last annotation update)			
DE	forkhead box protein f2 (forkhead-related protein FKHL5) (forkhead-related transcription factor 2) (FREC-2) (FREC-2)			
GN	FOXF2 OR FKHL5 OR FREAC2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99017977; PubMed=9799607;			
RA	Blitt A., Muhlappu M., Bjorseth C., Darnfors C., Johannesson T., Enerbaeck S., Carlsson P.;			
RT	"The two-exon gene of the human forkhead transcription factor FREAC-2. (FKHL5) is located at 6p25.3."			
RL	Genomics 53:387-390(1998).			
RN	[2]			
RP	SEQUENCE OF 37-444 FROM N.A.			
RP	TISSUE=Lung;			
RX	MEDLINE=96224034; PubMed=8626802;			
RA	Hellyqvist M., Muhlappu M., Samuelsson L., Enerbaeck S., Carlsson P.;			
RT	"Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2."			
RL	J. Biol. Chem. 271:4482-4490(1996).			
RN	[3]			
RP	PARTIAL SEQUENCE FROM N.A.			
RX	MEDLINE=95045392; PubMed=7957066;			
RA	Pierrou S., Hellyqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;			
RT	"Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending."			
RL	EMBO J. 13:5002-5012(1994).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98389768; PubMed=9722567;			
RA	Hellyqvist M., Muhlappu M., Blitt A., Enerbaeck S., Carlsson P.;			
RT	"The human forkhead protein FREAC-2 contains two functionally			



RT Redundant activation domains and interacts with TBP and TFIIB.";  
 RL J. Biol. Chem. 273:23335-23343(1998).  
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION ACTIVATOR FOR A NUMBER OF LUNG-  
 CC SPECIFIC GENES.  
 CC -1- SUBUNIT: INTERACTS WITH THE TRANSCRIPTION FACTORS TBP AND TFIIB.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: LUNG AND PLACENTA.  
 CC -1- DOMAIN: TWO ACTIVATION DOMAINS, AD1 AND AD2, C-TERMINAL OF (AND  
 CC DISTINCT FROM) THE FORKHEAD DOMAINS ARE NECESSARY FOR  
 CC TRANSCRIPTIONAL ACTIVATION.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF084939; AADI9875.1; .  
 DR EMBL: AF084938; AADI9875.1; JOINED.  
 DR EMBL: U13220; AAC32226.1; .  
 DR HSSP: 063245; 2HPH.  
 DR TRANSFAC: T02465; .  
 DR MIM: 603250; .  
 DR InterPro: IPR001766; Fork\_head.  
 DR Pfam: PF00250; Fork\_head; 1.  
 DR PRINTS: PRO0053; FORKHEAD.  
 DR SMART: SMO0339; FH; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation; Activator.  
 FT DOMAIN 33 41 POLY-ALA.  
 FT DOMAIN 46 55 POLY-SER.  
 FT DOMAIN 58 67 POLY-SER.  
 FT DOMAIN 77 84 POLY-GLY.  
 FT DNA\_BIND 99 190 FORK-HEAD.  
 FT DOMAIN 262 272 POLY-HIS.  
 FT DOMAIN 301 306 POLY-GLY.  
 FT DOMAIN 312 315 POLY-GLY.  
 FT DOMAIN 429 432 POLY-SER.  
 FT DOMAIN 444 459 POLY-HIS.  
 SQ SEQUENCE 444 AA; 45993 MW; 32BDC5F373CFB147 CRC64;  
 Query Match 6.5%; Score 96.5; DB 1; Length 444;  
 Best Local Similarity 28.4%; Pred. No. 5.9; Mismatches 63; Indels 27; Gaps 6;  
 Matches 42; Conservative 16; Mismatches 63; Indels 27; Gaps 6;  
 QY 115 TREHVPPRPKRKASHYPOKASKNPVSQVSTAFPTAATOLDGYPRAESSILTKS 174  
 DB 2 TTEGPPAPPLRRKSP-----VPGALQALMSPPPAAAAAAAPPTSSSSSS 53  
 QY 175 GSSCPYVSSVNHRTIPSIDASFVEKDDGCPPIETG--NNCSSSTSSPTTPPCSET- 231  
 DB 54 ASCASSSSSSNSASAPs---AACKSAGGAGAGSGAGAKKASSGLRR--PEKPYSYIA 107  
 QY 232 -----PEKYPDSQYKFIGSVF 250  
 DB 108 LIYMAIQSSPSK-RLTLSEIYQFLQARF 134  
 RESULT 13  
 DUP3\_RAT STANDARD; PRT; 977 AA.  
 AC P97838;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Disks large-associated protein 3 (DAP-3) (SAP90/PSD-95-associated  
 DE protein 3) (SAPAP3) (PSD-95/SAP90 binding protein 3).  
 GN DLAGP3 OR DAP3.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=brain;  
 RX MEDLINE=97277335; PubMed=9115257;  
 RA Takuchl M., Hata Y., Hirao K., Toyoda A., Irie M., Takai Y.;  
 RT "SAPAPs. A family of PSD-95/SAP90-associated proteins localized at  
 RT postsynaptic density.";  
 RL J. Biol. Chem. 272:11943-11951(1997).  
 CC -1- FUNCTION: May play a role in the molecular organization of  
 CC synapses and neuronal cell signaling. Could be an adapter protein  
 CC linking ion channel to the subynaptic cytoskeleton. May induce  
 CC enrichment of PSD-95/SAP90 at the plasma membrane.  
 CC -1- SUBUNIT: Interacts with Dlg1 and Dlg4/PSD-95 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed in brain.  
 CC -1- SIMILARITY: BELONGS TO THE SAPAP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U67139; AAB48589.1; .  
 KW Membrane.  
 FT DOMAIN 222 233 POLY-HIS.  
 SQ SEQUENCE 977 AA; 106019 MW; 1EFF13F5718C74FD CRC64;  
 Query Match 6.3%; Score 94.5; DB 1; Length 977;  
 Best Local Similarity 21.4%; Pred. No. 21; Mismatches 119; Indels 79; Gaps 12;  
 Matches 64; Conservative 37; Mismatches 119; Indels 79; Gaps 12;  
 QY 7 TYRD-SFETTSGGSSVDLVG--MALPLGLAPNLSSASVSASA-----SEDSA 49  
 DB 293 SYRDLSEFKRSGSE---GRLACTGMSMSLDGVSXKSAHNTIMVSGRGDYGAGCG 348  
 QY 50 KTRKRPYTTTKSRKESSEDEHDKFLDALQLFDRDKKIKIAYGSKTVIOIRHAKYFLK 109  
 DB 349 KGLIGPETHAKART-----YHYIQVPODDMGVPT-GGKDELPCRMRSGSYIK 397  
 QY 110 V---QKNGTREHVPPRPKRKASHYPOKASKNPVSQ----- 144  
 DB 398 AMGDESSGSDGSPKTSPPALARR-FASRRSSVDTARINCCVPRIRPRSSIPGYSNL 456  
 QY 145 ---QVSTAFPTAATYQDSCGYPPRAESSILTKSGSSCPYVSSVNHRTIPSIDAFYEKDD 201  
 DB 457 TTCQLSEFNQOLEAVGCVFGLSEQAVDALDLPCFCFMRs--HSYLAIAIGACSQDDDD 514  
 QY 202 GGP----PGIETGNCCSSTESSPTTPPCSEIPEKYPKPDPSQYKFIGSVFSDSTTD 256  
 DB 515 CLPILAAPASVSGRPGSSFNFRKAPPPIPPGSGAPPRI-----SITAGSSTD 561  
 RESULT 14  
 PAX6\_BRARE STANDARD; PRT; 437 AA.  
 ID PAX6\_BRARE  
 AC P26630;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Paired box protein PAX1(ZF-A) (PAX-6).  
 GN PAX6 OR PAXZF-A OR PAX1(ZF-A).  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Danio.

```

OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92037521; PubMed=1718739;
RA Krauss S., Johansen T., Korch V., Moens U., Ericson J.U., Fjose A.;
RT "zebrafish pax1(zf:a), a paired box-containing gene expressed in the
RL neural tube."
RN EMBL J. 10:3609-3619(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=91375540; PubMed=1660220;
RA Krauss S., Johansen T., Korch V., Fjose A.;
RT "Expression pattern of zebrafish pax genes suggests a role in early
RL brain regionalization."
RN Nature 353:267-270(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315909; PubMed=1352238;
RA Puschel A.W.P., Gruss P., Westerfield M.;
RT "Sequence and expression pattern of pax-6 are highly conserved
RL between zebrafish and mice."
RT Development 114:643-651(1992).
CC -1- FUNCTION: TRANSCRIPTION FACTOR EXPRESSED IN SPATIALLY RESTRICTED
CC REGIONS OF THE NEURAL TUBE DURING EMBRYONIC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM AND A SHORT FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SPATIALLY RESTRICTED REGIONS OF THE NEURAL
CC TUBE.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X61389; CAA43661.1; -
DR EMBL: X63183; CAA44867.1; -
DR PIR: S18558; S18558.
DR HSSP: P26367; 6PAX.
DR TRANSFAC: T00682; -
DR ZFIN: ZDB-GENE-990415-200; pax6a.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR001523; Paired_box.
DR Pfam: PF00046; homeobox.1.
DR Pfam: PF00292; PAX.1.
DR PRINTS: PR00027; PAIREDBOX.
DR SMART: SM00389; HOK; 1.
DR SMART: SM00351; PAX; 1.
DR PROSITE: PS00027; HOMEBOX.1; 1.
DR PROSITE: PS00071; HOMEBOX.2; 1.
DR PROSITE: PS00034; PAIRED_BOX.1.
DR Transcription regulation; Paired box; Homeobox; DNA-binding;
KW Nuclear protein; Developmental protein; Alternative splicing.
FT DOMAIN 23 147 PAIRED BOX.
FT DNA_BIND 229 288 HOMEBOX.
FT DOMAIN 298 437 SER/THR/PRO-RICH.
FT VARSPIC 66 66 Q --> OTHADAKYVDNEN (YN LONG ISOFORM).
FT SEQUENCE 437 AA; 46425 MW; 01D0999EFC22837 CRC64;

```

```

Query Match 6.3%; Score 94; DB 1; Length 437;
Best Local Similarity 19.9%; Pred. No. 8.7;
Matches 55; Conservative 54; Mismatches 91; Indels 76; Gaps 14;

```

```

OY 9 RDSFETTSGGSSVDLGNALDGLAPNLSSASVSASASDSKAKIRKPYTTKRSRNSQ 68
DB 193 QDGCQSDSGGE-----NTNNTSSNGEDSDETOMRLQLRKRKLQNRNTSTYOE 239

```

```

OY 69 EHDKFEALQLDRDKKI--EAFVGSKTVLQI---RSHAKYFLKYNKNGREHPP 122
DB 240 Q-----IEAL-----EKEPERHYPDVFRERLRKAKIDLPERRIDVWPSNRKARRRE-KLR 291
OY 123 RRRKASHDPKQASKNVVSOOVSTAF-----PTAATOLDSGYPRASSIITKSGS 176
DB 292 NGRQASN-----SSSHIRISSFSSTSVQPIPOPTTPPSFTSGSM-LGRSDTALTNTYS 345
OY 177 SCPTVSSW-----VHHTPSIDASFVEKDDGPPGTEGNN-----CS 214
DB 346 ALRPMPSFTMANNLPMQPSQTSYSCLPTSPSVNGR--SYDTYPPHQAHMNSQMAA 403
OY 215 SGSTES-----SPTTPCSEIPEKVKPDPSQOY 243
DB 404 SGTSTGLISPGVSVVQVPGSE-----PDMSQYM 433

RESULT 15
OTX3_BRARE
ID OTX3_BRARE STANDARD; PRT; 338 AA.
AC 090267;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Homeobox protein OTX3 (ZOTX3).
GN OTX3.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON AB;
RX MEDLINE=95206106; PubMed=7698305;
RA Mori H., Miyazaki Y., Morita T., Nitta H., Mishina M.;
RT "different spatio-temporal expressions of three otx homeoprotein
RL transcripts during zebrafish embryogenesis."
RL Brain Res. Mol. Brain Res. 27:221-231(1994).
CC -1- FUNCTION: MAY PLAY A ROLE IN VERY EARLY EMBRYOGENESIS,
CC GASTRULATION, AND THE DEVELOPMENT AND SUBDIVISION OF THE
CC Diencephalon AND THE MIDBRAIN. MAY PLAY A ROLE IN THE ORGANIZER
CC FUNCTION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- DEVELOPMENTAL STAGE: DETECTED ALONG THE MARGINAL ZONE OF SYMMETRIC
CC EMBRYOS AT 5 HRS OF DEVELOPMENT AND FOUND AT THE SHIELD, A PRIMARY
CC MORPHOLOGICAL ASYMMETRY AT 6 HRS OF DEVELOPMENT. FOUND IN TWO
CC STRIPES AT THE POSTERIOR SIDE OF THE EYE rudiments AT 12 HRS OF
CC DEVELOPMENT, DISTRIBUTED IN THE Diencephalon, MIDBRAIN AND THE
CC EPIPHYSIS AT 18 HRS OF DEVELOPMENT, AND FOUND IN THE Diencephalon
CC AND THE MIDBRAIN AT 24 HRS OF DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.
CC "BICOID" SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D26174; BAA05160.1; -
DR HSSP: P06601; 1FTJL.
DR ZFIN: ZDB-GENE-980526-27; otx3.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR003025; Otx_TF.
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00024; HOMEBOX.
DR PRINTS: PR01255; OTXHOMEBOX.
DR SMART: SM00389; HOK; 1.
DR PROSITE: PS00027; HOMEBOX.1; 1.

```

DR PROSITE; PS50071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
FT DNA\_BIND 38 97 HOMEBOX.  
FT DOMAIN 95 99 POLY-GLN.  
FT DOMAIN 134 144 POLY-SER.  
FT DOMAIN 158 162 POLY-SER.  
SQ SEQUENCE 338 AA; 36145 MW; ACFR6C2N43A69014 CRC64;

Query Match 6.2%; Score 93; DB 1; Length 338;

Best Local Similarity 19.9%; Pred. No. 7.6; Mismatches 103; Indels 50; Gaps 9;

Matches 48; Conservative 40; Mismatches 103; Indels 50; Gaps 9;

QY 25 GMALPGIAPNLSASVSASASEDSAKIRKPYTTKSGRESWSEQEHDKFLEALQLFDRDW 84  
DB 14 GIGIGGATMDLHPGSGVPTNPKRRRTTFRTO-----LDILESFAKT 60  
QY 85 KRTAFVGSKTIVIOI---RSHAQKF-----LKYQKNGTREHVPPPRKRKASHYP 133  
DB 61 RYPIEMKEEVALKINLPESHVQVWFKNRRKACRQOQOQSSSTNSKIRPAKKPSPSRESP 120  
QY 134 QKASKNPVPSQOVSTAFPTATQIDSGVYPRAESS-SILTKSGSSCPTVSSWVHTIPSI 192  
DB 121 GSESSGHFTPPAVSS--SSSSSSSSGNSALSGVGLISSSSGCTVPSPIMSPAFYSPV 178  
QY 193 DASFEKDDGGPGCI-----ETGNNC-----SSGSTESSPTWPPCSEIPEKVKPDPFSQV 242  
DB 179 PA-----PPIPLDISPPASASCMQRAMSSGGTGTGVPSYP---MPYNOAPSYAQG 225  
QY 243 Y 243  
DB 226 Y 226

Search completed: August 12, 2002, 10:12:32  
Job time: 225 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 12, 2002, 10:08:27 ; Search time 26.58 Seconds  
(without alignments)  
1902.077 Million cell updates/sec

Title: US-09-640-211A-1076  
Perfect score: 1492  
Sequence: 1 MPMLAEYRDSFETTSGCCS.....EMIQILIKLCTHEEPHNL 282

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 539846 seqs, 179280859 residues

Total number of hits satisfying chosen parameters: 539846

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/prodata/1/paa/PCOT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/prodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	701	47.0	336	6	US-10-155-881-21296
2	700	46.9	318	6	US-10-155-881-9544
3	700	46.9	325	6	US-10-155-881-20394
4	675.5	45.3	277	6	US-10-155-881-32157
5	664	44.5	304	6	US-10-155-881-32154
6	664	44.5	321	6	US-10-155-881-21324
7	628	42.1	313	6	US-10-155-881-28543
8	621	41.6	260	6	US-10-155-881-32135
9	618	41.4	333	6	US-10-155-881-28437
10	609.5	40.9	270	6	US-10-155-881-21319
11	585.5	39.2	231	6	US-10-155-881-10664
12	581	38.9	329	6	US-10-155-881-32140
13	578.5	38.8	187	6	US-10-155-881-9536
14	574.5	38.5	293	6	US-10-155-881-32148
15	550.5	36.9	214	6	US-10-138-145-1187
16	520.5	34.9	271	6	US-10-155-881-32151
17	520.5	34.9	359	6	US-10-155-881-9550
18	520	34.9	214	6	US-10-155-881-25530
19	519.5	34.8	229	6	US-10-155-881-32155
20	492	33.0	167	6	US-10-155-881-20389
21	489	32.8	181	6	US-10-155-881-21297
22	488.5	32.7	178	6	US-10-155-881-21320
23	481.5	32.3	174	6	US-10-155-881-21468
24	462	31.0	216	6	US-10-155-881-32691
25	460	30.8	207	6	US-10-155-881-10666
26	370.5	24.8	630	6	US-10-155-881-28422

27	367.5	24.6	267	6	US-10-155-881-18237	Sequence 18237, A
28	363	24.3	237	6	US-10-155-881-28014	Sequence 28014, A
29	363	24.3	260	6	US-10-155-881-28015	Sequence 28015, A
30	341	22.9	377	6	US-10-155-881-32144	Sequence 32144, A
31	339	22.7	169	6	US-10-155-881-28546	Sequence 28546, A
32	332	22.3	664	5	US-09-573-655B-1809	Sequence 1809, Ap
33	332	22.3	725	6	US-10-155-881-10033	Sequence 10033, A
34	331	22.2	438	6	US-10-155-881-10698	Sequence 10698, A
35	324	21.7	382	6	US-10-155-881-21365	Sequence 21365, A
36	324	21.7	387	5	US-09-935-625-5230	Sequence 5230, Ap
37	324	21.7	387	5	US-09-935-625-7742	Sequence 7742, Ap
38	324	21.7	481	6	US-10-155-881-28417	Sequence 28417, A
39	323.5	21.7	287	6	US-10-155-881-21366	Sequence 21366, A
40	323	21.6	557	6	US-10-155-881-25695	Sequence 25695, A
41	321	21.5	723	6	US-10-155-881-10030	Sequence 10030, A
42	320.5	21.5	745	6	US-10-155-881-26940	Sequence 26940, A
43	320	21.4	432	6	US-10-155-881-10671	Sequence 10671, A
44	312	20.9	287	5	US-09-935-625-1011	Sequence 1011, Ap
45	312	20.9	287	5	US-09-935-625-8409	Sequence 8409, Ap

## ALIGNMENTS

RESULT 1	US-10-155-881-21296	
;	Sequence 21296, Application US/10155881	
;	GENERAL INFORMATION:	
;	APPLICANT: Dolson, Stanton B.	
;	APPLICANT: Kovalic, David K.	
;	APPLICANT: Liu, Jingdong	
;	APPLICANT: Lufliya, Linda L.	
;	APPLICANT: McIninch, James	
;	TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH	
;	FILE REFERENCE: 38-21(15300)J	
;	CURRENT APPLICATION NUMBER: US/10/155,881	
;	CURRENT FILING DATE: 2002-05-22	
;	NUMBER OF SEQ ID NOS: 37595	
;	SEQ ID NO 21296	
;	LENGTH: 336	
;	TYPE: PRT	
;	ORGANISM: Glycine max	
;	US-10-155-881-21296	
Query Match	47.0%; Score 701; DB 6; Length 336;	
Best Local Similarity	58.6%; Pred. No. 5, 2e-46;	
Matches 153; Conservative	24; Mismatches 58; Indels 26; Gaps 7;	
QY 22	DLVGMALPG---LAPNLSSASVSASSEDASAKIRKPYRTTSRESGSEODHKLFAIQ 76	
DB 20	DPSNMPLPGVNNLPPPPPPAPAAVEDPNKTKIRPYRTTSRESWTQDHOKFLFAIQ 79	
QY 79	LFPDRKKIKLFAFGSKTVIIOIRSHAQKFLKYQKNGCTREHVPPRRKRASHYPOKASK 138	
DB 80	LFPDRKKIKLFAFGSKTVIIOIRSHAQKFLKYQKNGCTSEHVPPRRKRASHYPOKASK 139	
QY 139	NVPVSO-----QVSTAPPTAATOLDGSGYVPRASSSILTKSGSSCPTVSGWHHTTPST- 192	
DB 140	TPTVSOVGMGLQSSSAF-----IEPAYISPPSSSVLGPVYNNMP-LSSMWNNTTPQPG 192	
QY 193	DASFVEKDDCGPPPGIETG---NNCSSGSTSSSPPTPSPCEI-----PEYKPPQFSOV 242	
DB 193	NVPVYTDMDGLTGAGQADALNCCYSSSNBSPPTWPRKRINOQDQKPTKMPDFAOV 252	
QY 243	YKFIGSVFDPSTIDHLKLKE 263	
DB 253	YSFIGSVFDPSTNNHLKLQD 273	
RESULT 2	US-10-155-881-9544	

```
: Sequence 9544, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 9544
: LENGTH: 318
: TYPE: PRT
: ORGANISM: Zea mays
US-10-155-881-9544
```

```
Query Match 46.9%; Score 700; DB 6; Length 318;
Best Local Similarity 59.0%; Pred. No. 5.8e-46;
Matches 148; Conservative 26; Mismatches 45; Indels 32; Gaps 7;
```

```
OY 21 VDLYGMALPGLAPNLSASVSASSEDGAKIRKPYTTITKRESWSEQEHDKFLFALQLF 80
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 LSNVSASAP--PPPOAGSDAGSGEAEKSKVRKPYTTITKRESWTEQEHDKFLFALQLF 91
OY 81 DRDMKIAFVGSKTVLQIRSHAKQKFLVYOKNGTREHPPRRKRAHPYPOKASKNV 140
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 DRDMKIEAFVGSKTVLQIRSHAKQKFLVYOKNGTSEHPPRRKRAHPYPOKASKNE 151
OY 141 PVSQGVSTAFPTAATQLDSGYVPRAESSILTKSGSCPTVSSVHHHTIPSIDASFEVKD 200
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 P-----NYGLKTDSSSIHRNSGMV-SVSSWPHRSIPQAVASSMWKD 192
OY 201 DG-GPGIETGNCCSGSTESSPTWP-----PCSEIPE-KVPPDSQVYKFGVFPDP 252
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 LGAGTGP--PNNFCSSSTEGPPTWPGETNGPFIQLPDLMPDRAGVSYFLGSVFPD 249
OY 253 STDLHLKLKE 263
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 STDLHLKLKE 260
```

```
RESULT 3
US-10-155-881-20394
: Sequence 20394, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 20394
: LENGTH: 325
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-20394
```

```
Query Match 46.9%; Score 700; DB 6; Length 325;
Best Local Similarity 57.9%; Pred. No. 6e-46;
Matches 143; Conservative 33; Mismatches 49; Indels 20; Gaps 5;
OY 25 GMAIPGLAPNLS-SASVSASSEDGAKIRKPYTTITKRESWSEQEHDKFLFALQLFDRD 83
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 43 GMLPGNPPPAVSAAATAANSPEDDAKKIRKPYTTITKRENNTEPEHDKFLFALQLFDRD 102
OY 84 WKRIEAFVGSKTVLQIRSHAKQKFLVYOKNGTREHPPRRKRAHPYPOKASKNVPS 143
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 WKRIEAFVGSKTVLQIRSHAKQKFLVYOKNGTSEHPPRRKRAHPYPOKASKNAPVL 162
OY 144 QOVSTAFPTAATQLDSGYVPRAESSILTKSGSCPTVSSVHHHTIPSIDASFEVKDDG 203
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 163 SOVSGRFOSSALIEPGYILKNDAPPML-KTPIIMNTVASSMNNITQTANLS----- 213
OY 204 PPGIETGNCCSGSTESSPTWPCCSE-----IPEKVPDSQVYKFGVFPDPSTTD 256
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 -PYTKVNNPCSSG--ESTPVMVPGESNGGKNKIHLPLVLDFTQYVGFISVFDNATE 270
OY 257 HLKLKE 263
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 HLKLKK 277
```

```
RESULT 4
US-10-155-881-32157
: Sequence 32157, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 32157
: LENGTH: 277
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-32157
```

```
Query Match 45.3%; Score 675.5; DB 6; Length 277;
Best Local Similarity 58.1%; Pred. No. 3.7e-44;
Matches 136; Conservative 34; Mismatches 45; Indels 19; Gaps 4;
```

```
OY 37 SASVSASSEDGAKIRKPYTTITKRESWSEQEHDKFLFALQLFDRDMKIEAFVGSKTV 96
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 SAATAANSPEDDAKKIRKPYTTITKRENNTEPEHDKFLFALQLFDRDMKIEAFVGSKSV 67
OY 97 LQIRSHAKQKFLVYOKNGTREHPPRRKRAHPYPOKASKNVPSQOVSTAFPTAATD 156
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 LQIRSHAKQKFLVYOKNGTSEHPPRRKRAHPYPOKASKNAPVL SOVSGRFOSSAL 127
OY 157 LDSGYVPRAESSILTKSGSCPTVSSVHHHTIPSIDASFEVKDDGPPGIETGNCCSSG 216
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 LEPGYILKNDAPPML-KTPIIMNTVASSMNNITQTANLS-----PYTKVNNPCSSG 177
OY 217 STSSPEPTWPCSE-----IPEKVPDSQVYKFGVFPDPSTDLKLKE 263
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 -ESTPVMVPGESNGGKNKIHLPLVLDFTQYVGFISVFDNATEHLKLKK 229
```

```
RESULT 5
US-10-155-881-32154
: Sequence 32154, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 32154
: LENGTH: 277
: TYPE: PRT
: ORGANISM: Glycine max
US-10-155-881-32154
```

```

?
?
? FILE REFERENCE: 38-21(15300)J
? CURRENT APPLICATION NUMBER: US/10/155,881
? CURRENT FILING DATE: 2002-05-22
? NUMBER OF SEQ ID NOS: 37555
? SEQ ID NO: 32154
?
? LENGTH: 304
?
? TYPE: PRT
? ORGANISM: Glycine max
?
? US-10-155-881-32154

```

Query Match	44.5%	Score 664	DB 6	Length 304
Best Local	51.3%	Pred. NO. 3.2e-43		
Matches 138	Conservative 42	Mismatches 35	Indels 34	Gaps 5

[illegible]

```

RESULT 6
US-10-155-881-21324
Sequence 21324: Application US/10155881
GENERAL INFORMATION:
APPLICANT: Dolson, Stanton B.
APPLICANT: Kovalic, David K.
APPLICANT: Liu, Jingtong
APPLICANT: Iutifiya, Linda L.
APPLICANT: McIninch, James
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
FILE REFERENCE: 38-21(15300)J
CURRENT APPLICATION NUMBER: US/10/155,881
CURRENT FILING DATE: 2002-05-22
NUMBER OF SEQ. ID NOS: 37595
SEQ ID NO 21324
LENGTH: 321
TYPE: PRT
ORGANISM: Glycine max
US-10-155-881-21324

```

[illegible]

```

Oy 201 DGGPGIGTGNCSGSGTSESSPEPTMPCSEI-----PEKXKPPSSOYKRTGS 248
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 -----QKVNCCSSMS-----PAAOLVGESNGORRNSHPLRVLPDAEYVSTIGS 255
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 249 VFDPSTDLKKL-----EMIQILIK 270
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 256 VFDPVNTGHVOKLKKEMDPIDETLYLLMR 284
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-10-155-881-28543
: Sequence 28543, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dolson, Stanton B.
: APPLICANT: Kovacic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lufeiya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37555
: SEQ ID NO 28543
: LENGTH: 313
: TYPE: PRT
: ORGANISM: Oryza sativa
US-10-155-881-28543

```

```

Query Match          42.1%  Score 628;  DB 6;  Length 313;
Best Local Similarity 48.9%;  Pred. No. 1.9e-40;
Matches 139;  Conservative 19;  Mismatches 40;  Indels 86;  Gaps 6

OY 33 PNLSSASVASASASEDSAKKIRRPYTTTKSRESWSEDEHDKFLEALQL----- 79
      | : : : | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 8 PPDAAAAAAGSAGEDASKRRKRPYTTTKSRESWSEDEHDKFLEALQINHHNYEMLPYWS 67

OY 80 -----PDDDKKIEAFPGSKVIVIOIRSHAOQYPLK 109
      |||||
Db 68 PYLHMANNVFMRTKNTATVAVAVEFKLYTSFEDDMKRIEAFPGSKVIVIOIRSHAOQYPLK 127

OY 110 YOKNGTIREHVPPRPKRKAASHRPPOKASKKNVPYSGQVSTAFTPAATOLDGGYPRAES 169
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 YOKNGSEHVPPRPKRKAASHRPPOKASKKNRP-----GYTIKADSS 169

OY 170 IITKSGSS-----PTVSSWWHTIIPSIDSAFVEKDDGPGGIEGNGSSGSS 219
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 170 MLRELREHCHCYMDOLFPTNCSLIH-----GEFVYARDLSA--GAANPNFCSNSTE 221

OY 220 SSPPTWPCSEIPEKVKPDPFSQVYKFGVSEFDPSTTDHLKLKE 263
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 222 GPAAAMOP-----VMPDFAQYYSFISGVFPDSTSGHLOKLKE 258

```

```

RESULT      8
US-10-155-881-32135
; Sequence 32135, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingtong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(1530)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 32135

```

LENGTH: 260  
TYPE: PR1  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)-(260)  
OTHER INFORMATION: unsure at all xaa locations  
US-10-155-881-32135

Query Match 41.6%; Score 621; DB 6; Length 260;  
Best Local Similarity 55.6%; Pred. No. 5, 2e-40;  
Matches 125; Conservative 29; Mismatches 61; Indels 10; Gaps 3;

QY 46 EDSAKIRKPYTTIKSRSESEODEHDKLEALQDFDRMKIEAFVSGKTVQIRSHAK 105  
DB 2 EDQKRVKRPYTTIKSRSESEODEHDKLEALQDFDRMKIEAFVSGKTVQIRSHAK 61  
QY 106 YFLKQKNGTRENHPPRRPKRASHPRQKASKNVPVSOQVSTAFPLATQDSCGYPPRA 165  
DB 62 YFMKQKNGTSENHPPRRPKRASHPRQKASKNVLTVSOVARPLQSSSALSSSHYRP 121  
QY 166 ESSSILTKSSGSCPTVSSVWHTTIPSIDASFVEKDDGPGIETGNCCSGSTESSPPTW 225  
DB 122 DSSSVVTPVRQCGX-PSKGTNTPLVSLRVTKD--MYLMSQQLNPFSSSMENTPRGH 178  
QY 226 PCCSEI-----PEKVPDFSOVYKFGVFPDSTTHLKLKE 263  
DB 179 PISKOTDGDGKPTIVMPFAQYSPFGIVTDPNAINHLORLKO 223

RESULT 9  
US-10-155-881-28437  
Sequence 28437, Application US/10155881  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
FILE REFERENCE: 38-21(15300)J  
CURRENT APPLICATION NUMBER: US/10/155,881  
CURRENT FILING DATE: 2002-05-22  
NUMBER OF SEQ ID NOS: 37595  
SEQ ID NO 28437  
LENGTH: 333  
TYPE: PR1  
ORGANISM: Oryza sativa  
US-10-155-881-28437

Query Match 41.4%; Score 618; DB 6; Length 333;  
Best Local Similarity 50.0%; Pred. No. 1, 2e-39;  
Matches 136; Conservative 37; Mismatches 63; Indels 36; Gaps 10;

QY 17 GGSVDLVGALPGLADNLSSASVSDSAKIRKPYTTIKSRSESEODEHDKLEA 76  
DB 33 GKKEKQVVAAP-LQPPMAVPAAPAAVGE-ARKVKRPYTTIKSRSESEODEHDKLEA 90  
QY 77 LQFLDRMKIEAFVSGKTVQIRSHAKQFLKQKNGTSENHPPRRPKRASHPRQKA 136  
DB 91 LQFLDRMKIEAFVSGKTVQIRSHAKQFLKQKNGTSENHPPRRPKRASHPRQKA 150  
QY 137 SKNVPVSOQVSTAFPLATQDSCGYPPRAESSILTKSSGSCPTVSSVWHTTIPS 192  
DB 151 SKNVS-----PAISQPPPLGEGCCVMSDTSPIVINTNSA-VVPSMNSTAQPL 200  
QY 193 DASFEVDGPGIETGNCCSGSTESSPPTWPCSEI-----PEKVPDFSOVYK 245  
DB 201 SASHTQ---GTGAVAT-NMCCS-SIESPSTWPTSEAVQENMLRLRAMDPFAQYSP 254

QY 246 IGSVDPSTTHLKLK-----EWIQLILK 270  
DB 255 LGSIFDPDTSCHLQTLKAMPDPIVETVLLMR 286

RESULT 10  
US-10-155-881-21319  
Sequence 21319, Application US/10155881  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
FILE REFERENCE: 38-21(15300)J  
CURRENT APPLICATION NUMBER: US/10/155,881  
CURRENT FILING DATE: 2002-05-22  
NUMBER OF SEQ ID NOS: 37595  
SEQ ID NO 21319  
LENGTH: 270  
TYPE: PR1  
ORGANISM: Glycine max  
US-10-155-881-21319

Query Match 40.9%; Score 609.5; DB 6; Length 270;  
Best Local Similarity 55.0%; Pred. No. 4, 2e-39;  
Matches 132; Conservative 27; Mismatches 58; Indels 23; Gaps 5;

QY 36 SSASVSASASEDSAKIRKPYTTIKSRSESEODEHDKLEALQDFDRMKIEAFVSGKT 95  
DB 14 SGMMAAASDGSCKKRRPYTTIKSRSESEODEHDKLEALQDFDRMKIEAFVSGKT 73  
QY 96 VOIRSHAKQFLKQKNGTRENHPPRRPKRASHPRQKASKNVPVSOQVSTAFPLAT 155  
DB 74 VOIRSHAKQFLKQKNGTRENHPPRRPKRASHPRQKASKNVLTVSOVARPLQSSSALSSSHYRP 121  
QY 156 QLDGYPRAESSILTKSGS---SCPTVSSVWHTTIPSIDASFVEKDDGPGIETGN 211  
DB 134 TLARG-FAMWDESLMNGADKPMTCODELNNVHGG-----NEADISKGTQTT 183  
QY 212 NCSSGSTESSPPTWPCSEIPEKVK-----PDFSOVYKFGVFPDSTTHLKLKE 263  
DB 184 NSSLSGVGNSTRFL-LTSEIPKQKQAPVLHGLPDEAEVGFISVDFDETNDHVKLKE 242

RESULT 11  
US-10-155-881-10664  
Sequence 10664, Application US/10155881  
GENERAL INFORMATION:  
APPLICANT: Dotson, Stanton B.  
APPLICANT: Kovalic, David K.  
APPLICANT: Liu, Jingdong  
APPLICANT: Lutfiyya, Linda L.  
APPLICANT: McIninch, James  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
FILE REFERENCE: 38-21(15300)J  
CURRENT APPLICATION NUMBER: US/10/155,881  
CURRENT FILING DATE: 2002-05-22  
NUMBER OF SEQ ID NOS: 37595  
SEQ ID NO 10664  
LENGTH: 231  
TYPE: PR1  
ORGANISM: Zea mays  
US-10-155-881-10664

Query Match 39.2%; Score 585.5; DB 6; Length 231;  
Best Local Similarity 53.3%; Pred. No. 2, 4e-37;  
Matches 130; Conservative 20; Mismatches 57; Indels 37; Gaps 7;



```

RESULT 12
US-10-155-881-32140
: Sequence 32140, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dolson, Stanton B.
: APPLICANT: Kovallc, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NOCLETIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ. ID NOS: 37595
: SEQ ID NO 32140
: LENGTH: 329
: TYPE: PRT
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)-(329)
: OTHER INFORMATION: unsure at all Xaa locations
US-10-155-881-32140

```

```

RESULT 13
US-10-155-881-9536
: Sequence 9536, Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jingdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: McIninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: NUMBER OF SEQ ID NOS: 2002-05-22
: SEQ ID NO 9536
: LENGTH: 187
: TYPE: PRT
: ORGANISM: Zea mays
US-10-155-881-9536

```

```

RESULT 14
US-10-155-881-32148
: Sequence 32148: Application US/10155881
: GENERAL INFORMATION:
: APPLICANT: Dotson, Stanton B.
: APPLICANT: Kovalic, David K.
: APPLICANT: Liu, Jiongdong
: APPLICANT: Lutfiyya, Linda L.
: APPLICANT: Mcninch, James
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
: TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
: FILE REFERENCE: 38-21(15300)J
: CURRENT APPLICATION NUMBER: US/10/155,881
: CURRENT FILING DATE: 2002-05-22
: NUMBER OF SEQ ID NOS: 37595
: SEQ ID NO 32148
: LENGTH: 293
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-10-155-881-32148

```

Query Match	38.5%	DB	6	Length	293:
Best Local Similarity	47.0%:	Pred	No. 2, 3e-36:		
Matches	133:	Conservative	17:	Mismatches	62:
				Indels	71:
				Gaps	6

  

QY	5	AETTRDSFETTGSGSVLDVGMALPGLAPNLSSASVSASASEDSARKIKRPITITKSRRS	64
DB	14	ATTSIDATATPT-----IATTEAGGEAPPEKKVKAATITTSRS	50



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

## OM protein - protein search, using sw model

Run on: August 12, 2002, 10:03:27 ; Search time 109.58 Seconds  
(without alignments)  
905.809 Million cell updates/sec

Title: US-09-640-211a-1076

Perfect score: 1492  
Sequence: 1 MPMLAEYRDSFETSGSS.....EWIQLIKLCTHEEPHNL 282

## Scoring table:

BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep:\*  
25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep:\*  
26: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1492	100.0	282	1	PCT-US00-06112-1076
2	1492	100.0	282	1	US-09-640-211a-1076
3	724.5	48.6	290	18	US-09-452-997-12
4	701	47.0	332	26	US-60-324-109-23746
5	700	46.9	325	26	US-60-312-544-10669
6	695	46.6	327	1	PCT-US00-06112-2276
7	695	46.6	327	16	US-09-266-513-346

	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45																																																																																																																																																													
	695	46.6	327	20	US-09-640-211a-2276	678	45.4	286	18	US-09-452-997-20	664	44.5	317	26	US-60-312-544-10279	645	43.2	331	19	US-09-595-329A-2123	645	43.2	331	19	US-09-595-329A-2123	630	42.2	356	19	US-09-595-329A-2124	625	41.9	290	21	US-09-708-427-53221	625	41.9	295	21	US-09-708-427-53220	622.5	41.7	334	26	US-60-356-051-2044	608.5	40.8	302	20	US-09-620-394B-3147	606.5	40.7	284	20	US-09-620-394B-3148	598.5	40.1	302	21	PCT-US00-31414-12	576	38.6	287	26	US-60-356-051-1779	574.5	38.5	293	17	PCT-US00-31414-44	574.5	38.5	293	17	US-09-394-519-50	574.5	38.5	293	26	US-60-312-544-7855	574.5	38.5	293	26	US-60-324-109-33179	574.5	38.0	307	26	US-60-356-051-2104	560.5	37.6	214	21	US-09-733-089-18848	560.5	37.6	214	22	US-09-816-660-18848	555	37.2	275	18	US-09-452-997-24	551	36.9	422	21	US-09-733-089-761	551	36.9	422	22	US-09-816-660-761	521	34.9	264	21	US-09-733-089-14811	521	34.9	264	22	US-09-816-660-14811	520.5	34.9	278	26	US-60-312-544-8983	519.5	34.8	229	26	US-60-324-109-27074	519.5	34.8	229	26	US-60-324-109-33107	519.5	34.8	229	26	US-09-733-089-14750	503.5	33.7	250	22	US-09-816-660-14750	503.5	33.7	250	22	US-09-816-660-14750	502.5	33.7	213	18	US-09-452-997-25	502.5	33.7	213	18	US-09-708-427-8611	500.5	33.5	195	21	US-09-708-427-8612	497	33.3	129	21	US-09-733-089-22573	497	33.3	129	22	US-09-816-660-22573	492	33.0	130	21	US-09-708-427-51138	492	33.0	165	21	US-09-708-427-51136	490	32.8	128	26	US-60-312-544-6419

## ALIGNMENTS

RESULT 1

PCT-US00-06112-1076

Sequence 1076, Application PC/TUS0006112

GENERAL INFORMATION:

APPLICANT: Wood, Marion

APPLICANT: Shenk, Michael A.

APPLICANT: McGrath, Annette

TITLE OF INVENTION: Compositions and methods for the

TITLE OF INVENTION: modification of plant gene transcription.

FILE REFERENCE: 11000.1021C1PCT

CURRENT APPLICATION NUMBER: PCT/US00/06112

CURRENT FILING DATE: 2000-03-13

NUMBER OF SEQ ID NOS: 2368

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 1076

LENGTH: 282

TYPE: PRT

ORGANISM: Pinus radiata

PCT-US00-06112-1076

Query Match 100.0%; Score 1492; DB 1; Length 282;

Best Local Similarity 100.0%; Pred. No. 6-3e-120;

Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 MPMLAEYRDSFETSGSSVDLVGNALPGLAPNLSASVSASDSAKKIRKPYITRK 60

DB 1 MPMLAEYRDSFETSGSSVDLVGNALPGLAPNLSASVSASDSAKKIRKPYITRK 60

```

QY 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVP 120
; SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVP 120
Db 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVP 120
QY 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPT 100
; PPRKRRASHHPYPOKASKNVPVSOQVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPT 100
Db 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPT 100
QY 181 VSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNNSSSGSTESSPTWPCSEIPEKVKPDFS 240
; VSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNNSSSGSTESSPTWPCSEIPEKVKPDFS 240
Db 181 VSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNNSSSGSTESSPTWPCSEIPEKVKPDFS 240
QY 241 QVYKFGSVDFPSTTDLKLLKLEWIIQILKLCCTHEEPFNL 282
; QVYKFGSVDFPSTTDLKLLKLEWIIQILKLCCTHEEPFNL 282
Db 241 QVYKFGSVDFPSTTDLKLLKLEWIIQILKLCCTHEEPFNL 282

```

```

RESULT 2
US-09-640-211a-1076
; Sequence 1076, Application US/09640211a
; GENERAL INFORMATION:
; APPLICANT: Wood, Marlon
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; FILE REFERENCE: 11000.1021CIU
; CURRENT APPLICATION NUMBER: US/09/640,211a
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1076
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211a-1076

```

```

Query Match 100.0%; Score 1492; DB 20; Length 282;
Best Local Similarity 100.0%; Pred. No. 6.3e-120;
Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPMLAETTRDSFETTSAGSSVDLGMALPGIAPRLSASVSASASEDSAKKIRPYITK 60
Db 1 MPMLAETTRDSFETTSAGSSVDLGMALPGIAPRLSASVSASASEDSAKKIRPYITK 60
QY 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVP 120
; SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVP 120
Db 61 SRESMSQEHDKFLEALQLEFDRDKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVP 120
QY 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPT 180
; PPRKRRASHHPYPOKASKNVPVSOQVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPT 180
Db 121 PPRKRRASHHPYPOKASKNVPVSOQVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPT 180
QY 181 VSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNNSSSGSTESSPTWPCSEIPEKVKPDFS 240
; VSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNNSSSGSTESSPTWPCSEIPEKVKPDFS 240
Db 181 VSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNNSSSGSTESSPTWPCSEIPEKVKPDFS 240
QY 241 QVYKFGSVDFPSTTDLKLLKLEWIIQILKLCCTHEEPFNL 282
; QVYKFGSVDFPSTTDLKLLKLEWIIQILKLCCTHEEPFNL 282
Db 241 QVYKFGSVDFPSTTDLKLLKLEWIIQILKLCCTHEEPFNL 282

```

```

RESULT 3
US-09-452-997-12
; Sequence 12, Application US/09452997
; GENERAL INFORMATION:
; APPLICANT: Meng, Zude
; APPLICANT: Orozco, Emil M.
; TITLE OF INVENTION: Plant Circadian Rhythm-Myb Homologs
; FILE REFERENCE: B81296 US NA
; CURRENT APPLICATION NUMBER: US/09/452,997

```

```

; CURRENT FILING DATE: 1999-12-02
; EARLIER APPLICATION NUMBER: 60/110,780
; EARLIER FILING DATE: 1998-12-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-452-997-12

```

```

Query Match 48.6%; Score 724.5; DB 18; Length 290;
Best Local Similarity 61.8%; Pred. No. 1.1e-53;
Matches 147; Conservative 27; Mismatches 37; Indels 27; Gaps 5;

```

```

QY 33 PNLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQLEFDRDKKIEAFVGS 92
; PNLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQLEFDRDKKIEAFVGS 92
Db 8 PNLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQLEFDRDKKIEAFVGS 92
QY 93 SKTVIQRSHAQKYFLKVGKNGTREHVPVPPRKRASHHPYPOKASKNVPVSOQVSTAFPT 152
; SKTVIQRSHAQKYFLKVGKNGTREHVPVPPRKRASHHPYPOKASKNVPVSOQVSTAFPT 152
Db 93 SKTVIQRSHAQKYFLKVGKNGTREHVPVPPRKRASHHPYPOKASKNVPVSOQVSTAFPT 152
QY 153 AATQDSCGYTPRAESSILITKSGSSCPTVSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNN 212
; AATQDSCGYTPRAESSILITKSGSSCPTVSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNN 212
Db 153 AATQDSCGYTPRAESSILITKSGSSCPTVSSWVHHTTIFSIDASFVEKDDGPGGIEGTGNN 212
QY 213 CSSGSTRESSPTWPCSEIPEKVKPDFS-11000.1021CIU 263
; CSSGSTRESSPTWPCSEIPEKVKPDFS-11000.1021CIU 263
Db 213 CSSGSTRESSPTWPCSEIPEKVKPDFS-11000.1021CIU 263

```

```

RESULT 4
US-60-324-109-23746
; Sequence 23746, Application US/60324109
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)B
; CURRENT APPLICATION NUMBER: US/60/324,109
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 33196
; SEQ ID NO 23746
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
US-60-324-109-23746

```

```

Query Match 47.0%; Score 701; DB 26; Length 332;
Best Local Similarity 58.6%; Pred. No. 1.5e-51;
Matches 153; Conservative 24; Mismatches 58; Indels 26; Gaps 7;

```

```

QY 22 DLVGMALPG---LAPLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQ 78
; DLVGMALPG---LAPLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQ 78
Db 22 DLVGMALPG---LAPLSSASVSASEDSAKKIRPYITTKRESMSQEHDKFLEALQ 78
QY 79 LFDRODKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVPVPPRKRASHHPYPOKASK 138
; LFDRODKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVPVPPRKRASHHPYPOKASK 138
Db 79 LFDRODKKIEAFVGSKTVIQIRSHAQKYFLKVGKNGTREHVPVPPRKRASHHPYPOKASK 138
QY 139 NNPVSO-----QVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPTVSSWVHHTTIFST- 192
; NNPVSO-----QVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPTVSSWVHHTTIFST- 192
Db 139 NNPVSO-----QVSTAFPTAATQDSCGYTPRAESSILITKSGSSCPTVSSWVHHTTIFST- 192
QY 136 TPTVSOVMGRLSSSAF-----IEPATYIYSPDSSVLCTPTTNP-LSMWYNTTPOG 188
; TPTVSOVMGRLSSSAF-----IEPATYIYSPDSSVLCTPTTNP-LSMWYNTTPOG 188
Db 136 TPTVSOVMGRLSSSAF-----IEPATYIYSPDSSVLCTPTTNP-LSMWYNTTPOG 188
QY 193 DASFVEKDDGPGGIEGTG---NNCSTGSTRSSPTWPCSEI-----PEKVKPDFSQV 242
; DASFVEKDDGPGGIEGTG---NNCSTGSTRSSPTWPCSEI-----PEKVKPDFSQV 242

```

Db 189 NVQVTRDDMGILGAGGAAPLNCYSSSSNESTPPTWPRSRKINOGDQKPIKWPDPACV 248  
QY 243 YKFGISYFDPSTDLKLKE 263  
Db 249 YSFGISYFDPSTNHLKLOO 269

RESULT 5  
US-60-312-544-10669  
; Sequence 10669, Application US/60312544  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Edgerton, Michael D  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jiongong  
; APPLICANT: Stehn, Joshua  
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-10(52726)A  
; CURRENT APPLICATION NUMBER: US/60/312, 544  
; CURRENT FILING DATE: 2001-08-15  
; NUMBER OF SEQ ID NOS: 10730  
; SEQ ID NO 10669  
; LENGTH: 325  
; TYPE: PR  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: uC-gmromlsoy104a03\_FLI  
US-60-312-544-10669

Query Match 46.9%; Score 700; DB 26; Length 325;  
Best Local Similarity 57.9%; Pred. No. 1.7e-51;  
Matches 143; Conservative 35; Mismatches 49; Indels 20; Gaps 5;  
QY 25 GMAIPLGAPLNL-SASVSASASESDAKIRKPYTTTKSRRESWSEQEHDKFLALQLFDRD 83  
Db 43 GMLPFGMPPRASVSAATNAASPEDAAKIRKPYTTTKSRRESWTEPHDKFLALQLFDRD 102  
QY 84 WKTEAFVSGKTVIQRSHAKYFLKVNKGNGREHVPPRRKASHPPQKASKNVPS 143  
Db 103 WKTEAFVSGKTVIQRSHAKYFLKVNKGNGREHVPPRRKASHPPQKASKNVPS 162  
QY 144 QQVTAFTATATQDSCGYVRAESSILTKSGSCPTVSSVNHHTIPSIDASEYENDG 203  
Db 163 SOVSGRFOSSALIEPGYILKNAPPL-KTPIIMNTVASSMSNLTQTANLS----- 213  
QY 204 PPGIETGNCCSGSTESSPTWPCSE-----IPEKVPDFSOYKFTIGSVFDPSTTD 256  
Db 214 -PYTKVNNPCSSG--ESTPKVMPVSGESGOGNKTHTPLRLVLPDFTQYVGFISVFPDPNATE 270  
QY 257 HLKRLKE 263  
Db 271 HLKRLK 277

RESULT 6  
PCT-US00-06112-2276  
; Sequence 2276, Application PC/TUS0006112  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Marion  
; APPLICANT: Shenk, Michael A.  
; APPLICANT: McGrath, Annette  
; APPLICANT: Glenn, Matthew  
; TITLE OF INVENTION: Compositions and methods for the  
; FILE REFERENCE: 11000.1021CIPCT  
; CURRENT APPLICATION NUMBER: PCT/US00/06112  
; CURRENT FILING DATE: 2000-03-13  
; NUMBER OF SEQ ID NOS: 2368  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2276

; LENGTH: 327  
; TYPE: PR  
; ORGANISM: Eucalyptus grandis  
PCT-US00-06112-2276

Query Match 46.6%; Score 695; DB 1; Length 327;  
Best Local Similarity 56.4%; Pred. No. 4.7e-51;  
Matches 145; Conservative 33; Mismatches 55; Indels 24; Gaps 7;

QY 34 NLSASVSASA-----SEDSAKIRKPYTTTKSRRESWSEQEHDKFLALQLFDRDKKTEA 89  
Db 26 NAGSAAEGCGAAPPVADSDSKVRKPYTTTKSRRESWTEPHDKFLALQLFDRDKKTEA 85  
QY 90 FVSGKTVIQRSHAKYFLKVNKGNGREHVPPRRKASHPPQKASKNVPSQOVSTA 149  
Db 86 FVSGKTVIQRSHAKYFLKVNKGNGREHVPPRRKASHPPQKASKNVPSQOVSTA 144  
QY 150 FPTAATQDSCGYVRAESSILTKSGSCPTVSSVNHHTIPSIDASEYENDG--GPGI 207  
Db 145 FOVSSAFLEPGHIVRPGDGLNNSRTV-AUSSMNSVPMASASQGTGDVIGSGP-- 201  
QY 208 ETGNCCSGSTESSPTWPCSEIPE-----KVAPDFSOYKFTIGSVFDPSTDLKL 260  
Db 202 VPSNCCNCCSSNDSTPRSWPNAQAIETPLDQKHLRVMPDFAQYVRFISGVFDPDAGHQLR 261  
QY 261 LK-----EWIOLILK 270  
Db 262 LKQMDPINLETVLIMK 278

RESULT 7  
US-09-266-513-346  
; Sequence 346, Application US/09266513  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Marion  
; APPLICANT: McGrath, Annette  
; APPLICANT: Shenk, M. Andrew  
; APPLICANT: Glenn, Matthew  
; TITLE OF INVENTION: Compositions Isolated from Plant Cells  
; TITLE OF INVENTION: and their use in the modification of plant cell  
; FILE REFERENCE: 11000/1021  
; CURRENT APPLICATION NUMBER: US/09/266, 513  
; CURRENT FILING DATE: 1999-03-11  
; NUMBER OF SEQ ID NOS: 438  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 346  
; LENGTH: 327  
; TYPE: PR  
; ORGANISM: Eucalyptus grandis  
US-09-266-513-346

Query Match 46.6%; Score 695; DB 16; Length 327;  
Best Local Similarity 56.4%; Pred. No. 4.7e-51;  
Matches 145; Conservative 33; Mismatches 55; Indels 24; Gaps 7;  
QY 34 NLSASVSASA-----SEDSAKIRKPYTTTKSRRESWSEQEHDKFLALQLFDRDKKTEA 89  
Db 26 NAGSAAEGCGAAPPVADSDSKVRKPYTTTKSRRESWTEPHDKFLALQLFDRDKKTEA 85  
QY 90 FVSGKTVIQRSHAKYFLKVNKGNGREHVPPRRKASHPPQKASKNVPSQOVSTA 149  
Db 86 FVSGKTVIQRSHAKYFLKVNKGNGREHVPPRRKASHPPQKASKNVPSQOVSTA 144  
QY 150 FPTAATQDSCGYVRAESSILTKSGSCPTVSSVNHHTIPSIDASEYENDG--GPGI 207  
Db 145 FOVSSAFLEPGHIVRPGDGLNNSRTV-AUSSMNSVPMASASQGTGDVIGSGP-- 201  
QY 208 ETGNCCSGSTESSPTWPCSEIPE-----KVAPDFSOYKFTIGSVFDPSTDLKL 260  
Db 202 VPSNCCNCCSSNDSTPRSWPNAQAIETPLDQKHLRVMPDFAQYVRFISGVFDPDAGHQLR 261





```

Db 120 KRKAHPYPOAHKNVOL--QVPGSEFSTSEPNPSPFMPRESSSMLMTS---PTTAAA 173
QY 185 VHAHTIPSIDASFVEKDGGCPGDIETGNSSGSGTSSPPTWPCSEIPE-----K 234
Db 174 APWTNNMOTISFTL-----PKGAGANNCCSS--SEENTP---RRRSNDARDHGVGHSR 225
QY 235 VKPDFSOVKYFISGVDPSTTDHLKIKE 263
Db 226 VLPDFAQVYGFISGVDPYASNHLOKLK 254

```

# RESULT 14 US-09-708-427-53221

```

; Sequence 53221, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708.427
; NUMBER OF SEQ. ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53221
; LENGTH: 290
; TYPE: PRT
; ORGANISM: zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..290
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..290
; OTHER INFORMATION: Ceres Seq. ID 1930758
US-09-708-427-53221

```

Query Match 41.9%; Score 625; DB 21; Length 290;  
Best Local Similarity 53.6%; Pred. No. 4.4e-45;  
Matches 135; Conservative 23; Mismatches 50; Indels 44; Gaps 8;

```

QY 26 MALGGLAPNLSASVSASASDSAKIRKPYTTIKSRSESEQHDKFLALQLFDRDMK 85
Db 1 MVAAPAAAP-----AAEEYRKVKRPYTTIKSRSESWTEPHDKFLALQLFDRDMK 50
QY 86 KIEAFVGSKTVIQRSHAKYFLVKQNGTREHVPKPKKASHPYPOKASKNVP--VSO 144
Db 51 KIEAFVGSKTVIQRSHAKYFLVKQNGTGEHLPPRPKKAHPYPOKASKNAPAVSQ 110
QY 145 QV-STAPPTAATQDLSGYYPRAESSILTRSGSCPTVSSVNHHT-----IPSIDASFE 198
Db 111 AILSQEQPTQREQ-----GSVMPMDTATVNTNANVAVPSMDNTLAQ 152
QY 199 KDDCG-PPGIETGNSSGSGTSSPPTWPCSEI-----PEVKPDFSOVKYFISGV 250
Db 153 PFSAGHVQGAATNNCCS--SMESPSGTWPTSEAVEQENNVPLRAMPDFAQVYNFLGSIF 211
QY 251 DPSTTDHLKIK 262
Db 212 DPOTSGHLOMLK 223

```

```

RESULT 15
US-09-708-427-53220
; Sequence 53220, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708.427
; CURRENT FILING DATE: 2000-11-09

```

```

; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53220
; LENGTH: 295
; TYPE: PRT
; ORGANISM: zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..295
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..295
; OTHER INFORMATION: Ceres Seq. ID 1930757
US-09-708-427-53220

```

Query Match 41.9%; Score 625; DB 21; Length 295;  
Best Local Similarity 54.3%; Pred. No. 4.5e-45;  
Matches 134; Conservative 24; Mismatches 55; Indels 34; Gaps 7;

```

QY 31 LAFNLSASVSASASDSAKIRKPYTTIKSRSESEQHDKFLALQLFDRDMKIEAF 90
Db 1 MAPPPVVPAAAPAAEEYRKVKRPYTTIKSRSESWTEPHDKFLALQLFDRDMKIEAF 60
QY 91 VGSKTVIQRSHAKYFLVKQNGTREHVPKPKKASHPYPOKASKNVP--VSOQV-ST 148
Db 61 VGSKTVIQRSHAKYFLVKQNGTGEHLPPRPKKAHPYPOKASKNAPAVSQAILSQ 120
QY 149 APTTAATQDLSGYYPRAESSILTRSGSCPTVSSVNHHT-----IPSIDASFVEKDGG 203
Db 121 EOPTQREQ-----GSVMPMDTATVNTNANVAVPSMDNTLAQEPSAG 162
QY 204 -PPGIETGNSSGSGTSSPPTWPCSEI-----PEVKPDFSOVKYFISGVDPSTY 255
Db 163 HVGGAATNNCCS--SMESPSGTWPTSEAVEQENNVPLRAMPDFAQVYNFLGSIFDPOTS 221
QY 256 DHLKIK 262
Db 222 CHLOMLK 228

```

Search completed: August 12, 2002, 10:10:41  
Job time: 434 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:03:02 : Search time 13.08 Seconds  
(without alignments)  
526.607 Million cell updates/sec

Title: US-09-640-211a-1076

Perfect score: 1492  
Sequence: 1 MPMLEAFYRDSFTSGSS.....EWIQLIKCTHPEFHNL 282

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCFUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	335	22.5	83	US-09-202-712-16	Sequence 16, Appl
2	322	21.6	645	US-09-202-712-2	Sequence 2, Appl
3	238	16.0	83	US-09-202-712-14	Sequence 14, Appl
4	200	13.4	57	US-09-202-712-13	Sequence 13, Appl
5	195	13.1	55	US-09-202-712-7	Sequence 7, Appl
6	107	7.2	1123	US-08-408-865-1	Sequence 1, Appl
7	100.5	6.7	1105	US-08-999-774A-2	Sequence 2, Appl
8	96	6.4	1958	US-07-945-283-2	Sequence 2, Appl
9	92.5	6.2	568	US-08-320-559-30	Sequence 30, Appl
10	92.5	6.2	568	US-08-545-860D-30	Sequence 30, Appl
11	92.5	6.2	568	PCT-US94-004496-30	Sequence 30, Appl
12	91.5	6.1	410	US-09-431-573-4	Sequence 4, Appl
13	91.5	6.1	456	US-08-819-013-1	Sequence 1, Appl
14	91.5	6.1	1005	US-08-935-450-2	Sequence 2, Appl
15	90.5	6.1	410	US-08-123-343A-5	Sequence 5, Appl
16	90.5	6.1	410	US-08-123-343A-7	Sequence 7, Appl
17	90	6.0	430	US-08-427-993B-7	Sequence 7, Appl
18	90	6.0	430	US-08-478-609A-7	Sequence 7, Appl
19	90	6.0	509	US-08-427-993B-1	Sequence 1, Appl
20	90	6.0	509	US-08-478-609A-1	Sequence 1, Appl
21	90	6.0	1185	US-09-041-886-23	Sequence 23, Appl
22	87.5	5.9	268	US-09-187-049-9	Sequence 9, Appl
23	87.5	5.9	1259	US-09-187-049-13	Sequence 13, Appl
24	87	5.8	369	US-08-773-870-4	Sequence 4, Appl
25	87	5.8	369	US-09-393-569-2	Sequence 2, Appl
26	86.5	5.8	410	US-09-431-573-5	Sequence 5, Appl
27	86.5	5.8	672	US-08-049-254-2	Sequence 2, Appl

28	86.5	5.8	672	1	US-08-472-934-2	Sequence 2, Appl
29	86.5	5.8	672	2	US-08-323-460A-2	Sequence 2, Appl
30	86.5	5.8	672	2	US-08-461-146C-2	Sequence 2, Appl
31	86.5	5.8	672	3	US-08-461-146C-2	Sequence 2, Appl
32	86.5	5.8	672	4	US-08-628-829-2	Sequence 2, Appl
33	86.5	5.8	793	4	US-09-588-256-10	Sequence 10, Appl
34	86.5	5.8	1302	4	US-09-423-890-2	Sequence 2, Appl
35	86.5	5.8	1493	4	US-09-423-890-8	Sequence 8, Appl
36	86.5	5.8	1593	4	US-08-628-829-4	Sequence 4, Appl
37	86	5.8	553	3	US-09-083-351-2	Sequence 2, Appl
38	86	5.8	553	4	US-09-083-352-2	Sequence 2, Appl
39	86	5.8	1627	1	US-07-665-792E-9	Sequence 9, Appl
40	86	5.8	2842	1	US-07-741-940-7	Sequence 7, Appl
41	86	5.8	2842	1	US-08-289-548A-7	Sequence 7, Appl
42	86	5.8	2842	1	US-08-452-654-7	Sequence 7, Appl
43	86	5.8	2843	1	US-07-741-940-2	Sequence 2, Appl
44	86	5.8	2843	1	US-08-289-548A-2	Sequence 2, Appl
45	86	5.8	2843	1	US-08-452-654-2	Sequence 2, Appl

#### ALIGNMENTS

```

RESULT 1
US-09-202-712-16
; Sequence 16, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffer, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-202-712-16

Query Match      22.5%: Score 335; DB 4; Length 83;
Best Local Similarity 80.0%: Pred. No. 2,9e+24;
Matches 64; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Cy 41 SASASEDSAKKRPVYITTSRESWSQEHDKFLDALQFLDRDWKKIEAFVSGKYIQIR 100
Db 3 TTEGGEAPKKVRYAYITTSRESWTEGHEHDKFLDALQFLDRDWKKIEDFGSKYIQIR 62
Cy 101 SHAKYFLKVKNGSTREHVP 120
Db 63 SHAKYFLKVKNGSTLAHP 82

RESULT 2
US-09-202-712-2
; Sequence 2, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffer, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23

```

```
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-202-712-2
```

```
Query Match          21.6%; Score 322; DB 4; Length 645;
Best Local Similarity 34.5%; Pred. No. 8.4e-22;
Matches 80; Conservative 35; Mismatches 73; Indels 44; Gaps 7;
```

```
OY 43 SASBDSAKIRKPYTTTSSRSMSOEHDKFLALQLFDRDWMKIEAFVSGKTYIOIRSH 102
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 5 TSGEELLAKARKPYTTTQORRWTEDEHERFLRLRYGRAMORIEEHIGTKTAVOIRSH 64
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 103 AOKYFLKVKNGKRE-----HVPKPKRKASHIPPQKASKNVPSQOVSTAFPT 152
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 65 AOKFTTKLEKAEVYGIPVCOALDIETPPRPKRKPNTPYPRKGNNGTSSQVSSA--- 121
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 153 AATQDGSYVRAESSLTITKSSGSCPTVSSWHHTIPSTP-ASFVEKDDGCPGCIETGN 211
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 122 -----KDALVSSAS-----SSQLNQAFLDLKMKPFSEKSTGTGKENQD--E 160
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 212 NCSSGTESSPTWPCSEIPEKVKPDPGQYKFGISVDEPSTDLHLKILKE 263
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 161 NCSGVSTVKNKTPFL--PTKQVSGDIETSKT-----STVDNAVQDVPKKNKD 203
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
RESULT 3
US-09-202-712-14
; Sequence 14, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffner, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; EARLIER FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-202-712-14
```

```
Query Match          16.0%; Score 238; DB 4; Length 83;
Best Local Similarity 52.6%; Pred. No. 3.3e-15;
Matches 41; Conservative 18; Mismatches 19; Indels 0; Gaps 0;
```

```
OY 43 SASBDSAKIRKPYTTTSSRSMSOEHDKFLALQLFDRDWMKIEAFVSGKTYIOIRSH 102
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 5 TSGEELLAKARKPYTTTQORRWTEDEHERFLRLRYGRAMORIEEHIGTKTAVOIRSH 64
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 103 AOKYFLKVKNGKREHVP 120
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 65 AOKFTTKLEKAEVYGIP 82
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
RESULT 4
US-09-202-712-13
; Sequence 13, Application US/09202712
; Patent No. 6265637
```

```
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffner, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-202-712-13
```

```
Query Match          13.4%; Score 200; DB 4; Length 57;
Best Local Similarity 60.7%; Pred. No. 6.9e-12;
Matches 34; Conservative 13; Mismatches 9; Indels 0; Gaps 0;
```

```
OY 57 TITKRSRMSOEHDKFLALQLFDRDWMKIEAFVSGKTYIOIRSHAKYFLKVK 112
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 TITKORRWTEDEHERFLRLRYGRAMORIEEHIGTKTAVOIRSHAKOKEFTKLE 56
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
RESULT 5
US-09-202-712-7
; Sequence 7, Application US/09202712
; Patent No. 6265637
; GENERAL INFORMATION:
; APPLICANT: Coupland, George M
; APPLICANT: Schaffner, Robert J
; TITLE OF INVENTION: Genetic control of flowering
; FILE REFERENCE: 620-55
; CURRENT APPLICATION NUMBER: US/09/202,712
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: PCT/GB97/01676
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: GB 9613132.1
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-202-712-7
```

```
Query Match          13.1%; Score 195; DB 4; Length 55;
Best Local Similarity 60.0%; Pred. No. 1.9e-11;
Matches 33; Conservative 13; Mismatches 9; Indels 0; Gaps 0;
```

```
OY 57 TITKRSRMSOEHDKFLALQLFDRDWMKIEAFVSGKTYIOIRSHAKYFLKVK 111
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 TITKORRWTEDEHERFLRLRYGRAMORIEEHIGTKTAVOIRSHAKOKEFTKLE 55
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
RESULT 6
US-09-408-865-1
; Sequence 1, Application US/09408865A
; Patent No. 6329171
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/408,865A
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
```

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1123  
TYPE: prt  
ORGANISM: Homo sapiens  
US-09-408-865-1

Query Match 7.2%; Score 107; DB 4; Length 1123;  
Best Local Similarity 27.3%; Pred. No. 0.22; Indels 56; Gaps 12;  
Matches 65; Conservative 25; Mismatches 92;

QY 36 SSASVSASASSEDGAKIRKPYT-----ITSRE-SWSEQEHDKFLALQLEFRDMKRTFAF 90  
DB 611 SSSSPHSASSDSTKAPQPRSGAHLCDQENCTAGSKRPSCGA--DSKTVLTKSP 668  
QY 91 VSKTVIQRIS-----HAQKFLKVKOK-----NCTREHVPPRRKRKASHPYOKASK 138  
DB 669 VLSNTTTEPASTMSPPAKKLALSAKASTLMRATGNDLRPPSPSSDLTH--PKKTSH 726  
QY 139 NVPVSOQVSTAPPTATQDLSGYPPRAESSILTKSGSCPTVSSVWHTIPSIDAFYE 198  
DB 727 PV-----VASTWVHRAAVS---PAPOSSRLQPPSPSPHTLS---STPK----- 767  
QY 199 KDCGPGCIETGNKSSGSTE-----SSPTWPCSEIPEKYPFSGYKFEIG 247  
DB 768 -----PPTSEPRSCSSISITRLQVNMEDLVSLPHQLEPASERPQ--SPSEKRRKTVG 818

RESULT 7  
US-08-999-774A-2  
Sequence 2, Application US/08999774A  
Patent No. 6274312  
GENERAL INFORMATION:  
APPLICANT: Gish, Kurt C.  
APPLICANT: Seghezzi, Wolfgang  
APPLICANT: Shanahan, Frances  
APPLICANT: Lees, Emma M.  
APPLICANT: McLanahan, Terrell K.  
TITLE OF INVENTION: Intracellular Regulatory Molecules;  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: DNAX Research Institute  
STREET: 901 California Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/999,774A  
FILING DATE: 10-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/032,818  
FILING DATE: 11-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ching, Edwin P.  
REGISTRATION NUMBER: 34,090  
REFERENCE/DOCKET NUMBER: DX0646  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650)852-9196  
TELEFAX: (650)496-1200  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1105 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-999-774A-2

Query Match 6.7%; Score 100.5; DB 4; Length 1105;  
Best Local Similarity 22.3%; Pred. No. 0.87;  
Matches 59; Conservative 31; Mismatches 88; Indels 87; Gaps 12;

QY 30 GLAP-NLSASVSASAS-----EDSAK-----IKKPTTIKSR-----ESWS 66  
DB 565 GIVPLHLRSPPQPAQOMLNPEKKKRPVDLQNGRLTDIYSKRTLLKSKASAGRWMT 624  
QY 67 EDEHDKFLALQLEFRDMKRTFAFSGSKTVIQRSHAKYFLKVOKNGTREHVPPRRKR 126  
DB 625 EDETLLEALEMRYADDMKVKSEHVGSKRQDCILH-----FLRL----- 664  
QY 127 KASHPYPOK-----ASKNVPVSOQVSTAPPTATQDLSGYPPRAESSILTKSGSSC 178  
DB 665 PLEDPYLENSDASLAPLAYQVPFSGSNPVMSTVAF--LASVDDPRVASA-----AKAA 718  
QY 179 PIVSSVWHTIP-----SIDASFYEKDDGPGIETGNKSSGSTE 219  
DB 719 LEFESVREEVPLELVEAHVKVQEAARASGRVDPTY-----GLE--SSCTAGTGP 767  
QY 220 SSPTWPCSEIPEKYPFSGYK 244  
DB 768 DEPEKLEGAEEKMEADPDGQOPEK 792

RESULT 8  
US-07-945-283-2  
Sequence 2, Application US/07945283  
Patent No. 5352596  
GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
APPLICANT: Mesley, Ronald D.  
TITLE OF INVENTION: Pseudotables Virus Deletion Mutants  
TITLE OF INVENTION: Involving The EPO and LTR Genes  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESS: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P.  
REGISTRATION NUMBER: 27976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 309-685-4011 ext.513  
TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1958 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-945-283-2

Query Match 6.4%; Score 96; DB 1; Length 1958;  
Best Local Similarity 26.8%; Pred. No. 5.1;  
Matches 41; Conservative 14; Mismatches 64; Indels 34; Gaps 7;







```

STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,343A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,957
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MTT-5907A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 61861-9540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-123-343A-5

```

[illegible]

Search completed: August 12, 2002, 10:08:45  
Job time: 343 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:02:37 : Search time 29.84 Seconds  
(without alignments)  
1049.692 Million cell updates/sec

Title: US-09-640-211a-1076  
Perfect score: 1492  
Sequence: 1 MPMLAEYRDSFETTSGGSS.....EMILILKLCCTHEEPFNL 282

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1492	100.0	282	21	AA833027
2	695	46.6	327	21	AA833321
3	631.5	42.3	333	21	AA836897
4	616.5	41.3	318	21	AA836898
5	608.5	40.8	302	21	AA834908
6	608	40.8	303	21	AA845426
7	606.5	40.7	284	21	AA844909
8	606	40.6	285	21	AA845427
9	598.5	40.1	284	21	AA836899
10	598.5	40.1	302	22	AA802543
11	574.5	38.5	293	22	AA802559

12	568.5	38.1	293	21	AA823940	Arabidopsis thalia
13	480.5	32.2	142	21	AA832797	Eucalyptus grandis
14	480.5	32.2	150	21	AA833245	Eucalyptus grandis
15	322	21.6	178	21	AA820007	Arabidopsis thalia
16	322	21.6	203	21	AA820006	Arabidopsis thalia
17	322	21.6	645	19	AA842084	Amino acid sequenc
18	322	21.6	645	22	AA802547	A. thaliana transc
19	322	21.6	645	22	AA801903	Arabidopsis thalia
20	322	21.6	645	22	AA801926	Arabidopsis thalia
21	312	20.9	287	22	AA801952	Arabidopsis thalia
22	310.5	20.8	836	21	AA842123	Arabidopsis thalia
23	310.5	20.8	856	21	AA842122	Arabidopsis thalia
24	310.5	20.8	909	21	AA842121	Arabidopsis thalia
25	307	20.6	163	21	AA820008	Arabidopsis thalia
26	305	20.4	608	19	AA879280	Arabidopsis phytoc
27	305	20.4	608	22	AA802563	A. thaliana transc
28	305	20.4	608	22	AA801889	Arabidopsis thalia
29	305	20.4	608	22	AA801911	Arabidopsis thalia
30	302	20.2	59	21	AA833396	Eucalyptus grandis
31	248.5	16.7	102	21	AA833278	Pinus radiata tran
32	239	16.0	47	21	AA833395	Eucalyptus grandis
33	207	13.9	145	21	AA833174	Eucalyptus grandis
34	168	11.3	48	21	AA833374	Pinus radiata tran
35	145	9.7	42	21	AA833375	Pinus radiata tran
36	144.5	9.7	309	21	AA833232	Eucalyptus grandis
37	144.5	9.7	370	21	AA832774	Arabidopsis thalia
38	137	9.2	387	21	AA849848	Arabidopsis thalia
39	136.5	9.1	203	21	AA846375	Arabidopsis thalia
40	136.5	9.1	215	21	AA846374	Arabidopsis thalia
41	133.5	8.9	275	21	AA808359	Arabidopsis thalia
42	133.5	8.9	278	21	AA808358	Arabidopsis thalia
43	133.5	8.9	314	21	AA808357	Arabidopsis thalia
44	133.5	8.9	322	21	AA849849	Arabidopsis thalia
45	133.5	8.9	384	21	AA819453	Arabidopsis thalia

## ALIGNMENTS

RESULT 1  
AA833027  
ID AA833027 standard; Protein; 282 AA.  
XX  
AC AA833027;  
XX  
DT 25-JAN-2001 (first entry)  
XX  
DE Pinus radiata transcription factor protein sequence #154.  
XX  
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
KW poplar; sweetgum; teak; mahogany; bzip; G-box binding factor;  
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;  
KW type 2 Cys2His2; CCAAT box element; MYB.  
XX  
OS Pinus radiata.  
XX  
PN WO200053724-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 09-MAR-2000; 2000MO-US06112.  
XX  
PR 11-MAR-1999; 99US-0266513.  
XX  
PR 18-AUG-1999; 99US-0149485.  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
PI Wood M, McGrath A, Shenk MA, Glenn M;  
XX WPI; 2000-579369/54.  
XX

PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT  
 XX  
 XX  
 PS Claim 8; Page 398; 747pp; English.  
 XX  
 XX The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
 CC transcription factor. The transcription factor may be used to produce a  
 CC plant having modified gene expression such as a woody plant e.g. a  
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
 CC to modify the activity of a polypeptide in a plant. The transcription  
 CC factors of the present invention are members from the following families  
 CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
 CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
 CC Cys2His2, CCAAT box elements and MYB.  
 CC  
 XX  
 XX Sequence 282 AA;

Query Match 100.0%; Score 1492; DB 21; Length 282;  
 Best Local Similarity 100.0%; Pred. No. 4,2e-129;  
 Matches 282; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MPMLAETRYROSFETSGSSVDLVGMALPGAPMLSSASVSASASEDSAKKIRKPYITRK 60  
 Db 1 mpmlaetryrdsfetstgssvdilvgmalpglapmlssasvsasasedsakkirkpytlrk 60  
 OY 61 SRESMSDEHDKFLDALQLEDRDMKTEAFVGSKTVTQIRSHAKQYFLKVKNGNTRHVP 120  
 Db 61 sresmsdehdkflaqlqldrdmkkleafvgsktvtqirshakqyflkvngntrhvp 120  
 OY 121 PPRKAKRASHYPPQKASKNVPSQOVSTAFPTATQIDSGYPRARESSITITKSGSCPT 180  
 Db 121 pprkkrashpyppqkasknvpsqovstafptaatqidsygypraessitltksgsscpt 180  
 OY 181 VSSVNHHTIPSIDASFVEKDDGPGGIEGTGNCSSGSTESSPPWPCSEIPEKVKKPDFS 240  
 Db 181 vssvnhhtipdsidfekddgpggiegtgncssgstessppwpcseipekvkpdfs 240  
 OY 241 QYVFRTGVSFDPSTTHLKRLKEMTQILILKCTHPEPHNL 282  
 Db 241 qyvfktgsvfdpstthlklkewtlqlklctcheephnl 282

## RESULT 2

AAB33321  
 ID AAB33321 standard; Protein; 327 AA.

AC AAB33321;

DT 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor protein sequence #419.

XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB.  
 XX  
 XX Eucalyptus grandis.

PN W0200053724-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000MO-US06112.

XX 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESIS RES 6 DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 XX  
 PI Wood M, McGrath A, Shenk MA, Glenn M;  
 DR WPI; 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT  
 XX  
 XX  
 PS Claim 8; Page 724; 747pp; English.

CC The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
 CC transcription factor. The transcription factor may be used to produce a  
 CC plant having modified gene expression such as a woody plant e.g. a  
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
 CC to modify the activity of a polypeptide in a plant. The transcription  
 CC factors of the present invention are members from the following families  
 CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
 CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
 CC Cys2His2, CCAAT box elements and MYB.  
 CC  
 XX  
 XX Sequence 327 AA;

Query Match 46.6%; Score 695; DB 21; Length 327;  
 Best Local Similarity 56.4%; Pred. No. 1.4e-55;  
 Matches 145; Conservative 33; Mismatches 55; Indels 24; Gaps 7;

OY 34 NLSASVSASAAAA--SEDSAKKIRKPYITRKRESMSDEHDKFLDALQLEDRDMKTEAF 89  
 Db 26 nagsaaeggaapaydsdkvkrkytltkreswtegehdflaahlfddwkkiaa 85  
 OY 90 FVCSKVIVQIRSHAKQYFLKVKNGNTRHVPVPRPKKASHYPPQKASKNVPSQOVSTAF 149  
 Db 86 fvsckvtdqirshakqyflkvngntrhvpvprprkrkaahpyqgkapk-apvsvgnpp 144  
 OY 150 FPPAATQLODSGYTPRAESSITLKSSGSCPTVSSWVHHTIPSIDASFVEKDDG--GPCI 207  
 Db 145 fpyssaflpqhltvrdpgsallgnrstsv-alswshnsvpamsaagtkdvysgpp-- 201  
 OY 208 ETGNMNCSSGSTESSPPWPCSEIPE-----KVKPDPSQVYKFTGSVFDPTDHLKK 260  
 Db 202 vpsncsssndstprspwpaqalepldqgkhlrvmpdtaqvyrftgsvfdpdaaghlqr 261  
 OY 261 LK-----EMTQILILK 270  
 Db 262 lkqmdpnlctvylmk 278

## RESULT 3

AAG36897  
 ID AAG36897 standard; Protein; 333 AA.

AC AAG36897;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 45282.

XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.  
XX PF 25-FEB-2000. 2000EP-0301439.  
XX 25-FEB-1999. 99US-0121825.  
PR 05-MAR-1999. 99US-0123180.  
PR 09-MAR-1999. 99US-0123548.  
PR 23-MAR-1999. 99US-0125788.  
PR 25-MAR-1999. 99US-0126764.  
PR 29-MAR-1999. 99US-0126785.  
PR 01-APR-1999. 99US-0127462.  
PR 06-APR-1999. 99US-0128234.  
PR 16-APR-1999. 99US-0129845.  
PR 19-APR-1999. 99US-0130077.  
PR 21-APR-1999. 99US-0130449.  
PR 23-APR-1999. 99US-0130510.  
PR 23-APR-1999. 99US-0130891.  
PR 28-APR-1999. 99US-0131449.  
PR 30-APR-1999. 99US-0132048.  
PR 30-APR-1999. 99US-0132407.  
PR 04-MAY-1999. 99US-0132484.  
PR 05-MAY-1999. 99US-0132485.  
PR 06-MAY-1999. 99US-0132486.  
PR 06-MAY-1999. 99US-0132487.  
PR 07-MAY-1999. 99US-0132863.  
PR 11-MAY-1999. 99US-0134256.  
PR 14-MAY-1999. 99US-0134218.  
PR 14-MAY-1999. 99US-0134219.  
PR 14-MAY-1999. 99US-0134221.  
PR 14-MAY-1999. 99US-0134370.  
PR 18-MAY-1999. 99US-0134768.  
PR 19-MAY-1999. 99US-0135124.  
PR 20-MAY-1999. 99US-0135124.  
PR 21-MAY-1999. 99US-0135353.  
PR 24-MAY-1999. 99US-0135629.  
PR 25-MAY-1999. 99US-0136021.  
PR 27-MAY-1999. 99US-0136392.  
PR 28-MAY-1999. 99US-0136782.  
PR 01-JUN-1999. 99US-0137222.  
PR 03-JUN-1999. 99US-0137528.  
PR 04-JUN-1999. 99US-0137502.  
PR 07-JUN-1999. 99US-0137724.  
PR 08-JUN-1999. 99US-0138094.  
PR 10-JUN-1999. 99US-0138540.  
PR 10-JUN-1999. 99US-0138847.  
PR 14-JUN-1999. 99US-0139119.  
PR 16-JUN-1999. 99US-0139452.  
PR 16-JUN-1999. 99US-0139453.  
PR 17-JUN-1999. 99US-0139492.  
PR 18-JUN-1999. 99US-0139454.  
PR 18-JUN-1999. 99US-0139455.  
PR 18-JUN-1999. 99US-0139456.  
PR 18-JUN-1999. 99US-0139457.  
PR 18-JUN-1999. 99US-0139458.  
PR 18-JUN-1999. 99US-0139459.  
PR 18-JUN-1999. 99US-0139460.  
PR 18-JUN-1999. 99US-0139461.  
PR 18-JUN-1999. 99US-0139462.  
PR 18-JUN-1999. 99US-0139463.  
PR 18-JUN-1999. 99US-0139750.  
PR 18-JUN-1999. 99US-0139763.  
PR 21-JUN-1999. 99US-0139817.  
PR 22-JUN-1999. 99US-0139899.  
PR 23-JUN-1999. 99US-0140353.  
PR 23-JUN-1999. 99US-0140354.  
PR 24-JUN-1999. 99US-0140695.  
PR 28-JUN-1999. 99US-0140823.  
PR 29-JUN-1999. 99US-0140991.  
PR 30-JUN-1999. 99US-0141287.  
PR 01-JUL-1999. 99US-0141842.  
PR 01-JUL-1999. 99US-0142154.  
PR 02-JUL-1999. 99US-0142055.

PR 06-JUL-1999. 99US-0142390.  
PR 08-JUL-1999. 99US-0142803.  
PR 09-JUL-1999. 99US-0142920.  
PR 12-JUL-1999. 99US-0142977.  
PR 13-JUL-1999. 99US-0143542.  
PR 14-JUL-1999. 99US-0143624.  
PR 15-JUL-1999. 99US-0144005.  
PR 16-JUL-1999. 99US-0144085.  
PR 16-JUL-1999. 99US-0144086.  
PR 19-JUL-1999. 99US-0144325.  
PR 19-JUL-1999. 99US-0144331.  
PR 19-JUL-1999. 99US-0144332.  
PR 19-JUL-1999. 99US-0144333.  
PR 19-JUL-1999. 99US-0144334.  
PR 19-JUL-1999. 99US-0144335.  
PR 20-JUL-1999. 99US-0144352.  
PR 20-JUL-1999. 99US-0144632.  
PR 20-JUL-1999. 99US-0144684.  
PR 21-JUL-1999. 99US-0144814.  
PR 21-JUL-1999. 99US-0145086.  
PR 21-JUL-1999. 99US-0145088.  
PR 22-JUL-1999. 99US-0145085.  
PR 22-JUL-1999. 99US-0145087.  
PR 22-JUL-1999. 99US-0145089.  
PR 22-JUL-1999. 99US-0145192.  
PR 23-JUL-1999. 99US-0145145.  
PR 23-JUL-1999. 99US-0145218.  
PR 23-JUL-1999. 99US-0145224.  
PR 26-JUL-1999. 99US-0145276.  
PR 27-JUL-1999. 99US-0145913.  
PR 27-JUL-1999. 99US-0145918.  
PR 27-JUL-1999. 99US-0145919.  
PR 28-JUL-1999. 99US-0145951.  
PR 02-AUG-1999. 99US-0146386.  
PR 02-AUG-1999. 99US-0146388.  
PR 02-AUG-1999. 99US-0146389.  
PR 03-AUG-1999. 99US-0147038.  
PR 04-AUG-1999. 99US-0147204.  
PR 04-AUG-1999. 99US-0147302.  
PR 05-AUG-1999. 99US-0147192.  
PR 05-AUG-1999. 99US-0147260.  
PR 06-AUG-1999. 99US-0147303.  
PR 06-AUG-1999. 99US-0147416.  
PR 09-AUG-1999. 99US-0147493.  
PR 09-AUG-1999. 99US-0147935.  
PR 10-AUG-1999. 99US-0148171.  
PR 11-AUG-1999. 99US-0148319.  
PR 12-AUG-1999. 99US-0148341.  
PR 13-AUG-1999. 99US-0148565.  
PR 13-AUG-1999. 99US-0148684.  
PR 16-AUG-1999. 99US-0149368.  
PR 17-AUG-1999. 99US-0149175.  
PR 18-AUG-1999. 99US-0149426.  
PR 20-AUG-1999. 99US-0149722.  
PR 20-AUG-1999. 99US-0149723.  
PR 20-AUG-1999. 99US-0149929.  
PR 23-AUG-1999. 99US-0149902.  
PR 23-AUG-1999. 99US-0149930.  
PR 25-AUG-1999. 99US-0150566.  
PR 26-AUG-1999. 99US-0150884.  
PR 27-AUG-1999. 99US-0151065.  
PR 27-AUG-1999. 99US-0151066.  
PR 27-AUG-1999. 99US-0151080.  
PR 30-AUG-1999. 99US-0151303.  
PR 31-AUG-1999. 99US-0151438.  
PR 01-SEP-1999. 99US-0151930.  
PR 07-SEP-1999. 99US-0152363.  
PR 10-SEP-1999. 99US-0153070.  
PR 13-SEP-1999. 99US-0153758.  
PR 15-SEP-1999. 99US-0154018.  
PR 16-SEP-1999. 99US-0154039.  
PR 20-SEP-1999. 99US-0154779.  
PR 22-SEP-1999. 99US-0155139.



PR	23-JUN-1999	9905-0140354
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140695
PR	28-JUN-1999	9905-0140823
PR	28-JUN-1999	9905-0140991
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142154
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0142920
PR	12-JUL-1999	9905-0142977
PR	13-JUL-1999	9905-0143324
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	15-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	19-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144353
PR	20-JUL-1999	9905-0144354
PR	20-JUL-1999	9905-0144355
PR	21-JUL-1999	9905-0144356
PR	21-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145087
PR	22-JUL-1999	9905-0145088
PR	22-JUL-1999	9905-0145089
PR	22-JUL-1999	9905-0145092
PR	22-JUL-1999	9905-0145113
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145218
PR	23-JUL-1999	9905-0145224
PR	23-JUL-1999	9905-0145276
PR	26-JUL-1999	9905-0145373
PR	27-JUL-1999	9905-0145519
PR	27-JUL-1999	9905-0145519
PR	28-JUL-1999	9905-0145551
PR	02-AUG-1999	9905-0146386
PR	02-AUG-1999	9905-0146387
PR	02-AUG-1999	9905-0146389
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147192
PR	05-AUG-1999	9905-0147260
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147365
PR	06-AUG-1999	9905-0147463
PR	09-AUG-1999	9905-0147935
PR	10-AUG-1999	9905-0148171
PR	11-AUG-1999	9905-0148319
PR	12-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148565
PR	13-AUG-1999	9905-0148684
PR	16-AUG-1999	9905-0149156
PR	17-AUG-1999	9905-0149175
PR	18-AUG-1999	9905-0149422
PR	20-AUG-1999	9905-0149426
PR	20-AUG-1999	9905-0149723
PR	20-AUG-1999	9905-0149929
PR	23-AUG-1999	9905-0149930
PR	25-AUG-1999	9905-0150566
PR	26-AUG-1999	9905-0150884
PR	27-AUG-1999	9905-0151065
PR	27-AUG-1999	9905-0151066
PR	27-AUG-1999	9905-0151080
PR	30-AUG-1999	9905-0151083

PR	31-MRG-1999;	99S05-0151930;
PR	01-SEP-1999;	99S05-0151930;
PR	07-SEP-1999;	99S05-0152363;
PR	10-SEP-1999;	99S05-0153070;
PR	13-SEP-1999;	99S05-0153758;
PR	15-SEP-1999;	99S05-0154019;
PR	16-SEP-1999;	99S05-0154038;
PR	20-SEP-1999;	99S05-0154179;
PR	22-SEP-1999;	99S05-0155139;
PR	23-SEP-1999;	99S05-0155486;
PR	24-SEP-1999;	99S05-0156559;
PR	28-SEP-1999;	99S05-0156458;
PR	29-SEP-1999;	99S05-0156566;
PR	04-OCT-1999;	99S05-0157117;
PR	04-OCT-1999;	99S05-0157153;
PR	05-OCT-1999;	99S05-0157753;
PR	06-OCT-1999;	99S05-0157865;
PR	07-OCT-1999;	99S05-0158029;
PR	08-OCT-1999;	99S05-0158133;
PR	12-OCT-1999;	99S05-0158316;
PR	13-OCT-1999;	99S05-0159293;
PR	13-OCT-1999;	99S05-0159294;
PR	13-OCT-1999;	99S05-0159325;
PR	14-OCT-1999;	99S05-0159329;
PR	14-OCT-1999;	99S05-0159330;
PR	14-OCT-1999;	99S05-0159331;
PR	14-OCT-1999;	99S05-0159637;
PR	14-OCT-1999;	99S05-0159684;
PR	18-OCT-1999;	99S05-0159684;
PR	21-OCT-1999;	99S05-0160741;
PR	21-OCT-1999;	99S05-0160767;
PR	21-OCT-1999;	99S05-0160768;
PR	21-OCT-1999;	99S05-0160770;
PR	21-OCT-1999;	99S05-0160814;
PR	21-OCT-1999;	99S05-0160815;
PR	22-OCT-1999;	99S05-0160815;
PR	22-OCT-1999;	99S05-0160981;
PR	22-OCT-1999;	99S05-0160981;
PR	23-OCT-1999;	99S05-0161004;
PR	23-OCT-1999;	99S05-0161005;
PR	23-OCT-1999;	99S05-0161405;
PR	26-OCT-1999;	99S05-0161359;
PR	26-OCT-1999;	99S05-0161360;
PR	26-OCT-1999;	99S05-0161361;
PR	28-OCT-1999;	99S05-0161920;
PR	28-OCT-1999;	99S05-0161920;
PR	28-OCT-1999;	99S05-0161993;
PR	28-OCT-1999;	99S05-0162143;
PR	28-OCT-1999;	99S05-0162144;

Best Local Similarity 51.4%; Pred. No. 2.4e-48;  
Matches 142; Conservative 30; Mismatches 55; Indels 49; Gaps 8;

[illegible]

```
RESULT      5
AA014908
ID  AA014908 standard; Protein: 302 AA.
XX
AC  AA014908;
XX
DT  17-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 14947.
XX
KW  Protein identification: signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
XX  termination sequence.
XX  Arabidopsis thaliana.
XX  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
XX  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999; 99US-0121825.
PR  05-MAR-1999; 99US-0123180.
PR  09-MAR-1999; 99US-0123548.
PR  23-MAR-1999; 99US-0125788.
PR  25-MAR-1999; 99US-0126264.
PR  29-MAR-1999; 99US-0126785.
PR  01-APR-1999; 99US-0127462.
PR  06-APR-1999; 99US-0128234.
PR  08-APR-1999; 99US-0128714.
PR  16-APR-1999; 99US-0129845.
PR  19-APR-1999; 99US-0130077.
PR  21-APR-1999; 99US-0130449.
PR  23-APR-1999; 99US-0130510.
PR  23-APR-1999; 99US-0130891.
PR  28-APR-1999; 99US-0131449.
PR  30-APR-1999; 99US-0132048.
PR  04-MAY-1999; 99US-0132407.
PR  05-MAY-1999; 99US-0132484.
PR  06-MAY-1999; 99US-0132485.
PR  06-MAY-1999; 99US-0132486.
PR  07-MAY-1999; 99US-0132487.
PR  11-MAY-1999; 99US-0132863.
PR  14-MAY-1999; 99US-0134256.
PR  14-MAY-1999; 99US-0134218.
PR  14-MAY-1999; 99US-0134219.
PR  14-MAY-1999; 99US-0134221.
PR  14-MAY-1999; 99US-0134370.
PR  18-MAY-1999; 99US-0134768.
PR  19-MAY-1999; 99US-0134941.
PR  20-MAY-1999; 99US-0135124.
PR  21-MAY-1999; 99US-0135353.
PR  24-MAY-1999; 99US-0135629.
PR  25-MAY-1999; 99US-0136021.
PR  27-MAY-1999; 99US-0136392.
PR  28-MAY-1999; 99US-0136782.
PR  01-JUN-1999; 99US-0137222.
PR  03-JUN-1999; 99US-0137528.
PR  04-JUN-1999; 99US-0137502.
PR  07-JUN-1999; 99US-0137724.
PR  08-JUN-1999; 99US-0138094.
PR  10-JUN-1999; 99US-0138540.
PR  10-JUN-1999; 99US-0138847.
PR  14-JUN-1999; 99US-0139119.
PR  16-JUN-1999; 99US-0139452.
PR  16-JUN-1999; 99US-0139453.
PR  17-JUN-1999; 99US-0139492.
PR  18-JUN-1999; 99US-0139454.
PR  18-JUN-1999; 99US-0139455.
PR  18-JUN-1999; 99US-0139456.
PR  18-JUN-1999; 99US-0139457.
PR  18-JUN-1999; 99US-0139458.
PR  18-JUN-1999; 99US-0139459.
PR  18-JUN-1999; 99US-0139460.
PR  18-JUN-1999; 99US-0139461.
PR  18-JUN-1999; 99US-0139462.
PR  18-JUN-1999; 99US-0139463.
PR  18-JUN-1999; 99US-0139750.
PR  18-JUN-1999; 99US-0139763.
PR  21-JUN-1999; 99US-0139817.
PR  22-JUN-1999; 99US-0139899.
PR  23-JUN-1999; 99US-0140353.
PR  23-JUN-1999; 99US-0140354.
PR  24-JUN-1999; 99US-0140695.
PR  28-JUN-1999; 99US-0140823.
PR  29-JUN-1999; 99US-0140991.
PR  30-JUN-1999; 99US-0141287.
PR  01-JUL-1999; 99US-0141842.
PR  01-JUL-1999; 99US-0142154.
PR  02-JUL-1999; 99US-0142055.
PR  06-JUL-1999; 99US-0142390.
PR  08-JUL-1999; 99US-0142803.
PR  09-JUL-1999; 99US-0142920.
PR  12-JUL-1999; 99US-0142977.
PR  13-JUL-1999; 99US-0143542.
PR  14-JUL-1999; 99US-0143624.
PR  15-JUL-1999; 99US-0144005.
PR  16-JUL-1999; 99US-0144085.
PR  16-JUL-1999; 99US-0144086.
PR  19-JUL-1999; 99US-0144325.
PR  19-JUL-1999; 99US-0144331.
PR  19-JUL-1999; 99US-0144332.
PR  19-JUL-1999; 99US-0144333.
PR  19-JUL-1999; 99US-0144334.
PR  19-JUL-1999; 99US-0144335.
PR  20-JUL-1999; 99US-0144352.
PR  20-JUL-1999; 99US-0144632.
PR  20-JUL-1999; 99US-0144684.
PR  21-JUL-1999; 99US-0144814.
PR  21-JUL-1999; 99US-0145086.
PR  21-JUL-1999; 99US-0145088.
PR  22-JUL-1999; 99US-0145085.
PR  22-JUL-1999; 99US-0145087.
PR  22-JUL-1999; 99US-0145089.
PR  22-JUL-1999; 99US-0145192.
PR  23-JUL-1999; 99US-0145195.
PR  23-JUL-1999; 99US-0145218.
PR  23-JUL-1999; 99US-0145224.
PR  26-JUL-1999; 99US-0145276.
PR  27-JUL-1999; 99US-0145913.
PR  27-JUL-1999; 99US-0145918.
PR  27-JUL-1999; 99US-0145919.
PR  28-JUL-1999; 99US-0145951.
PR  02-AUG-1999; 99US-0146386.
PR  02-AUG-1999; 99US-0146388.
PR  02-AUG-1999; 99US-0146389.
PR  03-AUG-1999; 99US-0147038.
PR  04-AUG-1999; 99US-0147204.
PR  04-AUG-1999; 99US-0147302.
PR  05-AUG-1999; 99US-0147192.
PR  05-AUG-1999; 99US-0147260.
PR  06-AUG-1999; 99US-0147303.
PR  06-AUG-1999; 99US-0147416.
PR  09-AUG-1999; 99US-0147493.
PR  09-AUG-1999; 99US-0147935.
PR  10-AUG-1999; 99US-0148171.
PR  11-AUG-1999; 99US-0148319.
PR  12-AUG-1999; 99US-0148341.
PR  13-AUG-1999; 99US-0148655.
PR  13-AUG-1999; 99US-0148684.
PR  16-AUG-1999; 99US-0149368.
PR  17-AUG-1999; 99US-0149175.
PR  18-AUG-1999; 99US-0149426.
PR  20-AUG-1999; 99US-0149722.
PR  20-AUG-1999; 99US-0149723.
```

PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158569.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 40.8%; Score 608.5; DB 21; Length 302;

Best Local Similarity 50.6%; Pred. No. 1.2e-47;

Matches 131; Conservative 34; Mismatches 65; Indels 29; Gaps 9;

QY 22 DIVGMALP-----GLAPNLSASVSASASEDSAKTIRKPYTIKSPREWSGEQHDKFLFA 76  
DB 15 dsnsmispsdsdgsfpaspatgrts- lvsfsecdptkikrpytlkksrenwctdghdkflfa 73  
QY 77 LOLEFDRMKRIEAFVSGSKTYIQIRSHAQKTYFLKVOKNGTREHYVPPPRKPKASHPYPOKA 136  
DB 74 lhlfdfrwkkkleafvsgsktvtgrtshaqkyflrvqksganehlppprpkrrkashpyrpika 133  
QY 137 SKNVPVSOQVSTAFPTAT--QLDSCGYTPRAESSILTKSGSCPYVSSVWHHTIRSIDA 194  
DB 134 pknv-----aytslpsstlpllepgylyssdsksimngnqavcastssw-nhestlnlpx 187

QY 195 SFEKDDGGPGPIET-----GNMCSSGSTE-----SSPTMPKCEIPEKVKPDPFSQYK 244  
DB 188 pyleee-----pvsataplpmrcrgdeteravatkpneesc-ekphtmpntaeyys 242  
QY 245 FIGSVFDPSTTDHLKLRKE 263  
DB 243 figsvfdpntsgldqlrkq 261  
RESULT 6  
AAC45426  
ID AAC45426 standard; Protein: 303 AA.  
XX  
AC AAC45426;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 57030.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142350.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143674.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144864.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161820.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Query Match

Best Local Similarity 40.8%; Score 608; DB 21; Length 303;  
Pred. No. 1.3e-47;

Matches 131; Conservative 34; Mismatches 66; Indels 28; Gaps 9;

OY 22 DVGMAIP-----GLAPNLSASVSASAPDSAKRTKRPYTTKRSRESSEOHDKFLRA 76  
DB 15 dssnmispsdsgfgisipatgrts-tvtfseadptkirkpytlkksrenwcdqendkfkica 73







PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145324.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158339.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161923.  
PR 29-OCT-1999; 99US-0162142.

Query Match	40.68;	Score 606;	DB 21;	Length 285;
Best Local Similarity	51.08;	Pred. No. 1.9e-47;		
Matches 130;	Conservative 34;	Mismatches 63;	Indels 28;	Gaps 9;

[illegible]

RESULT	9
AAAG36899	
ID	AAAG36899 standard; Protein: 284 AA.
XX	
AC	AAAG36899;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 45284.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130448.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.

PR	30-APR-1995	99US-0132048
PR	30-APR-1995	99US-0132407
PR	04-MAY-1995	99US-0132484
PR	05-MAY-1995	99US-0132485
PR	06-MAY-1995	99US-0132486
PR	06-MAY-1995	99US-0132487
PR	07-MAY-1995	99US-0132863
PR	11-MAY-1995	99US-0134256
PR	14-MAY-1995	99US-0134218
PR	14-MAY-1995	99US-0134219
PR	14-MAY-1995	99US-0134321
PR	14-MAY-1995	99US-0134370
PR	18-MAY-1995	99US-0134768
PR	19-MAY-1995	99US-0134411
PR	20-MAY-1995	99US-0135124
PR	21-MAY-1995	99US-0135513
PR	24-MAY-1995	99US-0135629
PR	25-MAY-1995	99US-0136021
PR	27-MAY-1995	99US-0136592
PR	28-MAY-1995	99US-0136782
PR	01-JUN-1995	99US-0137222
PR	03-JUN-1995	99US-0137528
PR	04-JUN-1995	99US-0137502
PR	07-JUN-1995	99US-0137724
PR	08-JUN-1995	99US-0138094
PR	10-JUN-1995	99US-0138540
PR	10-JUN-1995	99US-0138547
PR	14-JUN-1995	99US-0139152
PR	16-JUN-1995	99US-0139452
PR	16-JUN-1995	99US-0139453
PR	17-JUN-1995	99US-0139492
PR	18-JUN-1995	99US-0139454
PR	18-JUN-1995	99US-0139455
PR	18-JUN-1995	99US-0139456
PR	18-JUN-1995	99US-0139457
PR	18-JUN-1995	99US-0139458
PR	18-JUN-1995	99US-0139459
PR	18-JUN-1995	99US-0139460
PR	18-JUN-1995	99US-0139461
PR	18-JUN-1995	99US-0139462
PR	18-JUN-1995	99US-0139463
PR	18-JUN-1995	99US-0139750
PR	18-JUN-1995	99US-0139751
PR	18-JUN-1995	99US-0139763
PR	21-JUN-1995	99US-0139617
PR	22-JUN-1995	99US-0140091
PR	23-JUN-1995	99US-0140353
PR	23-JUN-1995	99US-0140354
PR	24-JUN-1995	99US-0140695
PR	28-JUN-1995	99US-0140823
PR	29-JUN-1995	99US-0140911
PR	30-JUN-1995	99US-0141287
PR	01-JUL-1995	99US-0141842
PR	01-JUL-1995	99US-0142154
PR	02-JUL-1995	99US-0142055
PR	06-JUL-1995	99US-0142390
PR	08-JUL-1995	99US-0142803
PR	09-JUL-1995	99US-0142820
PR	12-JUL-1995	99US-0142877
PR	13-JUL-1995	99US-0143342
PR	14-JUL-1995	99US-0143524
PR	15-JUL-1995	99US-0144005
PR	16-JUL-1995	99US-0144086
PR	16-JUL-1995	99US-0144086
PR	19-JUL-1995	99US-0144321
PR	19-JUL-1995	99US-0144332
PR	19-JUL-1995	99US-0144334
PR	19-JUL-1995	99US-0144334
PR	19-JUL-1995	99US-0144335
PR	20-JUL-1995	99US-0144352
PR	20-JUL-1995	99US-0144353
PR	20-JUL-1995	99US-0144884
PR	21-JUL-1995	99US-0144814

PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 28-JUL-1999; 99US-0145351.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-015138.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 08-OCT-1999; 99US-0158029.  
PR 12-OCT-1999; 99US-0158332.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.

PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 40.1%; Score 598.5; DB 21; Length 284;  
Best Local Similarity 55.6%; Pred. No. 9,1e-47;  
Matches 133; Conservative 26; Mismatches 53; Indels 27; Gaps 6;

QY 42 ASASEDSAKRKIRKPYITKRSRSMSRQEHDKFLFALQLFDRBWKRIEAFVSKTYIOTRS 101  
Db 3 sseedlskkrkpycltksreswlependkileaqlldrtwkkileaigsklvqirs 62  
QY 102 HAQYFLKQKNGTREVHPRPKRKASHPYOKASKNVPVSOQVSTAPPTATOLDSCY 161  
Db 63 haqyflkqksgtgehlpprpkrkaahpyqkakhkvql--qypgsktsesepndpsf 120  
QY 162 YPPAESSTILTKSGSCPVTSSKWHHTIPSIDASFVEKDDG-----GPPDIETGNMCS 214  
Db 121 mfpesssmfms---ptlaaapwtmnaqtisfcpjpkaycscstlmetagaganncs 176  
QY 215 SCGTSSPPTWPPCSEIPE-----KVKPDFOYVKFFIGSVFDPSTTHLKLKE 263  
Db 177 s--scentp---fprnrdardhgnvghslrvlpdlaqvygfyldpyashnlqklk 231

RESULT 10  
AAE02543  
ID AAE02543 standard; Protein; 302 AA.  
XX  
AC AAE02543;  
XX  
DT 10-AUG-2001 (first entry)  
XX  
DE A. thaliana transcription factor G254.  
XX  
KW plant transcription factor; phenotype; sugar sensing characteristic;  
KW transgenic plant; plant yield; growth; germination; photosynthesis;  
KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;  
KW storage organ; metabolism.  
XX  
OS Arabidopsis thaliana.  
OS  
FH Key Location/Qualifiers  
FT Domain 62..106  
FT /note="Conserved domain"  
PN WO200135725-A1.  
XX  
PD 25-MAY-2001.  
XX  
PF 14-NOV-2000; 2000MO-US31414.  
XX  
PR 17-NOV-1999; 99US-0166228.  
PR 17-APR-2000; 2000US-0197899.  
PR 22-AUG-2000; 2000US-0227439.  
XX

PA (MENDEL-) MENDEL, BIOTECHNOLOGY INC.  
 PA (JIAN/) JIANG C.  
 PA (HEAR/) HEARD J.  
 PA (PINE/) PINEDA O.  
 PA (PIIG/) PILGRIM M.  
 PA (ADAM/) ADAM L.  
 PA (RIEC/) RIECHMANN J. L.  
 PA (YUGG/) YU G.  
 PA (SAMA/) SAMAHA R.  
 XX  
 PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann J.L;  
 PI Yu G, Samaha R;  
 XX  
 DR WPI: 2001-335977/35.  
 DR N-PSDB: AAD06644.  
 XX  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 PT for altering the sugar sensing characteristics of plants and increasing  
 PT yield, e.g. corn, potato and cotton plants -  
 XX  
 PS Claim 4: Page 69-70; 151pp; English.

CC The patent relates to polynucleotides encoding 35 plant transcription  
 CC factors which may be used to modify phenotype associated with a plant's  
 CC sugar sensing characteristics and increasing yield when their expression  
 CC level is altered. Sugars are central regulatory molecules that control  
 CC aspects of physiology, metabolism and development. Therefore the cDNAs  
 CC and proteins of the invention are useful for modifying the growth and  
 CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
 CC respiration, starch and sucrose synthesis and degradation, pathogen  
 CC response, wounding response, cell cycle regulation, pigmentation,  
 CC flowering and senescence of plants and for modifying sink-source  
 CC relationships in seeds, tubers, roots, and other storage organs leading  
 CC to an increase in yield. The transcription factor polynucleotides and  
 CC polypeptides may be used to alter the structure and developmental  
 CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
 CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
 CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,  
 CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,  
 CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
 CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.  
 CC The present sequence is an Arabidopsis thaliana transcription factor.  
 CC  
 XX  
 SO Sequence 302 AA:

Query Match 40.1%; Score 598.5; DB 22; Length 302;  
 Best Local Similarity 50.2%; Pred. No. 1e-46;  
 Matches 130; Conservative 34; Mismatches 66; Indels 29; Gaps 9;

QY 22 DLVGMALP-----GLAPRLSSASVSASASEDSAKKIRKPTTITRSRSWSOEHDKFLA 76  
 DB 15 dasmslpsdsgfsglpatgrts-tvafsedpttkirkpytlkxstrnwtdqehdkllea 73  
 QY 77 LQLPDRMKRTKTEAFVSGKTVQIRSHOKYFLKYOKNGTEHVPKPPKPKASHPYOKA 136  
 DB 74 lhlfdtdkklkkaevgtaktvqirshqkyflkvyksganehlplpkkkashipyika 133  
 QY 137 SKNVVSOQVSTAFPTAAT--QLDSGYYPRAESSILTKSGSSCPYVSWVNHHTPSIDA 194  
 DB 134 pkay-----aytslpsatllplepylydsdskslmgndavcasassw-nhesnlpk 187  
 QY 195 SFVEKDDCGPCGIEI-----GNKSSSGSTE-----SSPTMPPCSEIPEKYPFSGYK 244  
 DB 188 pvlsee-----pvasataplpmnfrgdehtervavtkpneesc-ekphrvmpnafevys 242  
 QY 245 FICGVDPSTTDLKLKLE 263  
 DB 243 figsvfdpntsglqrlkq 261

RESULT 11  
 AAE02559

ID AAE02559 standard; Protein: 293 AA.  
 XX  
 AC AAE02559;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE A. thaliana transcription factor G254 homolog, G673.  
 XX  
 KW Plant transcription factor; phenotype; sugar sensing characteristic;  
 KW transgenic plant; plant yield; growth; germination; photosynthesis;  
 KW glyoxylate metabolism; respiration; pathogen response; wounding response;  
 KW cell cycle regulation; pigmentation; flowering; senescence; physiology;  
 KW storage organ; metabolism.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 37..95  
 FT /note= "Conserved domain"  
 XX  
 PN MO200135725-A1.  
 XX  
 PD 25-MAY-2001.  
 XX  
 PF 14-NOV-2000; 2000WO-US31414.  
 XX  
 PR 17-NOV-1999; 99US-0166228.  
 PR 17-APR-2000; 2000US-0197899.  
 PR 22-APR-2000; 2000US-0227439.  
 XX  
 XX (MENDEL-) MENDEL, BIOTECHNOLOGY INC.  
 PA (JIAN/) JIANG C.  
 PA (HEAR/) HEARD J.  
 PA (PINE/) PINEDA O.  
 PA (PIIG/) PILGRIM M.  
 PA (ADAM/) ADAM L.  
 PA (RIEC/) RIECHMANN J. L.  
 PA (YUGG/) YU G.  
 PA (SAMA/) SAMAHA R.  
 PA (SAMA/) SAMAHA R.  
 PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann J.L;  
 PI Yu G, Samaha R;  
 XX  
 DR WPI: 2001-335977/35.  
 DR N-PSDB: AAD06660.  
 XX  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 PT for altering the sugar sensing characteristics of plants and increasing  
 PT yield, e.g. corn, potato and cotton plants -  
 XX  
 PS Claim 4: Page 112-113; 151pp; English.

CC The patent relates to polynucleotides encoding 35 plant transcription  
 CC factors which may be used to modify phenotype associated with a plant's  
 CC sugar sensing characteristics and increasing yield when their expression  
 CC level is altered. Sugars are central regulatory molecules that control  
 CC aspects of physiology, metabolism and development. Therefore the cDNAs  
 CC and proteins of the invention are useful for modifying the growth and  
 CC germination rates of plants, photosynthesis, glyoxylate metabolism,  
 CC respiration, starch and sucrose synthesis and degradation, pathogen  
 CC response, wounding response, cell cycle regulation, pigmentation,  
 CC flowering and senescence of plants and for modifying sink-source  
 CC relationships in seeds, tubers, roots, and other storage organs leading  
 CC to an increase in yield. The transcription factor polynucleotides and  
 CC polypeptides may be used to alter the structure and developmental  
 CC characteristics of plants such as soybean, wheat, corn, potato, cotton,  
 CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,  
 CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,  
 CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,  
 CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,  
 CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.  
 CC The present sequence is a homolog of Arabidopsis thaliana transcription  
 CC factor.



```

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146366.
PR 02-AUG-1999; 99US-0146368.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```

```

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 38.1%; Score 568.5; DB 21; Length 293;
Best Local Similarity 46.6%; Pred. No. 5,5e-44;
Matches 132; Conservative 17; Mismatches 63; Indels 71; Gaps 6;

```

```

QY 5 AETFRDSFERTSGSSSDLVGMALPGLAPMLSSASVSASASEPSAKKIRKPYTTTSRES 64
DB 14 Aetstdelet-----latteageapekvrkayltksres 50
QY 65 WSEQEHDKPLEALQLPDRDKKITEAFVGSKTVIQRSHACKYFLVKYOKNCRHPPPPR 124
DB 51 wtgeghnkflleaqlfdrdkkkledfygsktvigishbakgyltkyqngtlhpypprp 110
QY 125 KRRASHPYPOKASKNVPVSOQVSTAPPATATOLDGYYPAESSILITKSGSSCPTVSSW 184
DB 111 krkaahypqkasknaqmslhvsmfptqlnlt-pytlpddtsallnlt-----avsgv 164
QY 185 VHHRTIPSIDASFVEKDGGCPGIEFGNCCS-----GSTESSPTWPPCSIPPKVKR---- 236
DB 165 I-----ppedelatlclgaevdvgsndmlsetspasgigssrllsd 206
QY 237 -----PDFSOVYKFGSVDPDPTDHLKKLKE 263
DB 207 skglrlakgapsmbglpdaevynflgsvfdpdksgymkklxe 249

RESULT 13
AAB32797
ID AAB32797 standard; protein; 142 AA.
XX
AC AAB32797;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor protein sequence #255.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MAUS;
KW homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
XX
XX type 2 Cys2His2; CCAT box element; MYB.
OS Eucalyptus grandis.
XX
PN MO200053724-A2.
XX
PD 14-SEP-2000.
XX
PE 09-MAR-2000; 2000MO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (PLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for

```



PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide  
PS  
XX  
PS Claim 8; Page 308; 747pp; English.

CC The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
CC Cys2His2, CCAAT box elements and MYB.

XX  
SQ Sequence 142 AA;

Query Match 32.2%; Score 480.5; DB 21; Length 142;  
Best Local Similarity 76.4%; Pred. No. 2,5e-36;  
Matches 94; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

OY 31 LAPNLSASVSASASDSAKRKRPYTIKSRRESMSEOHDKFLEALQLFDRDMKRIEAF 90

DB 12 matsttactcpa-agggdgqkvrkpytlktrsteeehndkflaqltdrdwkkiedf 70

OY 91 VGSRTVTOIRSHAKQYFLKYQKNGTREHVPPEPRKRASHPYPOKASKNVPSOOVSTAF 150

DB 71 vgsrtvlgtrshaqkyflkvqngavahvppprpkrtkaahpyqkasknvlvplqasmaq 130

OY 151 PTA 153

DB 131 pss 133

#### RESULT 14

AAB33245  
ID AAB33245 standard; Protein; 150 AA.

AC AAB33245;

DT 25-JAN-2001 (first entry)

DE Eucalyptus grandis transcription factor protein sequence #402.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;

KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;

KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;

KW type 2 Cys2His2; CCAAT box element; MYB.

XX Eucalyptus grandis.

OS

XX

PN WO200053724-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000MO-US06112.

PR 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;

PI WPI: 2000-579369/54.

XX New isolated polynucleotide encoding a plant transcription factor for

PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide  
PS  
XX  
PS Claim 8; Page 692; 747pp; English.

CC The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
CC zipper, LIM domain, AP2 and ERBs, zinc finger domains of type 2  
CC Cys2His2, CCAAT box elements and MYB.

XX  
SQ Sequence 150 AA;

Query Match 32.2%; Score 480.5; DB 21; Length 150;  
Best Local Similarity 76.4%; Pred. No. 2,7e-36;  
Matches 94; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

OY 31 LAPNLSASVSASASDSAKRKRPYTIKSRRESMSEOHDKFLEALQLFDRDMKRIEAF 90

DB 20 matsttactcpa-agggdgqkvrkpytlktrsteeehndkflaqltdrdwkkiedf 78

OY 91 VGSRTVTOIRSHAKQYFLKYQKNGTREHVPPEPRKRASHPYPOKASKNVPSOOVSTAF 150

DB 79 vgsrtvlgtrshaqkyflkvqngavahvppprpkrtkaahpyqkasknvlvplqasmaq 138

OY 151 PTA 153

DB 139 pss 141

#### RESULT 15

AAG20007  
ID AAG20007 standard; Protein; 178 AA.

AC AAG20007;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 22027.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

OS

XX

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129645.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137562.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145274.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.



•  
•  
•  
•

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2002, 01:49:35 ; Search time 1594.5 Seconds  
(without alignments)  
7296.566 Million cell updates/sec

Title: US-09-640-211A-2076  
Perfect score: 862  
Sequence: 1 caaacgctccgtctctc.....gcttaaaaaaaaaaaaaa 862

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estum:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	245.4	28.5	500	A1897784	A1897784 EST267227
2	245.4	28.5	592	AM928286	AM928286 EST307029
3	243.8	28.3	519	A1897681	A1897681 EST267124
4	242	28.1	476	BE057370	BE057370 sm02c04.y
5	241.8	28.1	395	BE805071	BE805071 ss37a07.y
6	241.8	28.1	523	A1930997	A1930997 sb45h07.y
7	238.8	27.7	425	AM423958	AM423958 sb58c05.y
8	238.2	27.6	688	B1272897	B1272897 NF091A12F
9	237.6	27.6	558	BE324639	BE324639 NF024C04P
10	237.6	27.6	669	BE457971	BE457971 NF037A10P
11	236.4	27.4	447	B1273011	B1273011 NF097F02F
12	235.8	27.4	681	A1486576	A1486576 EST244897
13	234.8	27.2	409	BE804790	BE804790 ss45f11.y
14	204.8	23.8	370	BE642441	BE642441 EST355917
15	203.2	23.6	371	A1487923	A1487923 EST246245
16	200.2	23.2	782	BE658316	BE658316 GM700005B
17	198.2	23.0	470	AM156542	AM156542 se28d11.y

18	193.4	22.4	727	10	BG607379	BG607379 WHE2471.H
19	189.8	22.0	626	9	AM255388	AM255388 M407 pep
20	188.6	21.9	718	10	BG369720	BG369720 HYSME1002
21	183.4	21.3	366	10	BM135365	BM135365 WHE0457.B
22	180	20.9	510	10	BF325282	BF325282 su02603.y
23	173.6	20.1	492	10	B1262794	B1262794 NF094E02E
24	170.8	19.8	320	9	A1489912	A1489912 EST248251
25	167.8	19.5	467	10	BE807621	BE807621 ss28c05.y
26	164.6	19.1	588	10	BE649523	BE649523 NF080G02E
27	164.2	19.0	557	10	BM178036	BM178036 sa168c09.
28	162	18.8	635	9	AM032656	AM032656 EST276215
29	161.4	18.7	412	9	AM756015	AM756015 s112e11.y
30	158.4	18.4	452	10	BF425079	BF425079 su41a07.y
31	158.4	18.4	536	10	B1701275	B1701275 sa956608
32	158.4	18.4	566	10	B1424418	B1424418 sa956608
33	157	18.2	408	10	BF647022	BF647022 NF026F03E
34	157	18.2	567	9	AM737355	AM737355 EST338878
35	156.8	18.2	686	10	BG320722	BG320722 Zm04_07a0
36	156	18.1	502	10	BE459188	BE459188 EST414480
37	155.4	18.0	734	9	AT055122	AT055122 com0003B
38	154.8	18.0	631	9	AL507948	AL507948 AL507948
39	152.6	17.7	649	9	BE124286	BE124286 EST394411
40	150.4	17.4	715	10	BE658737	BE658737 GM700007A
41	149.4	17.3	442	10	BM356958	BM356958 12111-F11
42	149.4	17.3	601	9	AM672062	AM672062 LG1.354.G
43	148.2	17.2	454	10	BG509767	BG509767 sa225f10
44	145.2	16.8	256	10	BE805198	BE805198 sa39a09.y
45	142.6	16.5	600	10	BE489186	BE489186 WHE1075.G

#### ALIGNMENTS

RESULT 1  
A1897784  
LOCUS EST267227 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
DEFINITION c18D30123, mRNA sequence.  
ACCESSION A1897784  
VERSION A1897784.2 GI:11388186  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids I: Solanales: Solanales: Solanaceae: Solanum: Lycopersicon.  
Alcala, J., Vrebalov, J., White, R., Matera, A. L., Vlasov, T., Holt, I. E., Liang, F., Upton, J., Roming, C. M., Craven, M. B., Fujii, C. Y., Bowman, C. L., Nieman, W., Fraser, C. M., Venter, J. C., Martin, G. B., Tanksley, S. D., and Giovannoni, J. Generation of ESTs from tomato carpel tissue. Unpublished (1999)  
On Jul 27, 1999 this sequence version replaced gi:5603686.  
Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
FEATURES  
source  
1..500  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="c18D30123"  
/clone\_jib="tomato ovary, TAMU"  
/tissue\_type="carpel"  
/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab\_host="XLI-Blue MRF"  
/note="Vector: pBluescript SK(-); Site-1: EcorI; Site-2: XhoI; c18D - Tomato Carpel EST library. OligodT-primed and

directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."

BASE COUNT 185 a 70 c 110 g 135 t

Query Match 28.5%; Score 245.4; DB 9; Length 500;  
Best Local Similarity 74.5%; Pred. No. 1.7e-44;  
Matches 309; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

ORIGIN

5 cgtctcgtttctctccaaagctgacatgagcaagaagccagagccagacagtgtaagt 64  
|||||  
18 CATCTTTTCTTTTAAAAAATATATAAAAAAGATGATGATTAATGCACT 77  
|||||  
65 cccaagatgtcgagtgagaaaagccgctgagacatgagaaagagatcatccatca 124  
|||||  
78 CTCAGATGTTGAAGTGAAGAAAGGCGCTTGACATGGAAGAAGATTATTTCTCATTA 137  
|||||  
125 actacatacgaaatcacagcggaagcagttggaactccctagccaaagctgctgtctaa 184  
|||||  
138 ACTCATTTGCTTAATCATGTGTAAGGTGTGTAACCTCTAGCTAAATCTGCTGCTCA 197  
|||||  
185 aaagtaacgggaagagtgctgcgtccgctgagactctctgcgacccgagctccgga 244  
|||||  
198 AACCTACTGGAAGAAAGTTGTAAGACTCCGATGGCTAAATTTCTTGACCTGATGTGCA 257  
|||||  
245 gaagcaacatcacactactgaggaagcaagctcctgatactgagacatgagcaagctggagaa 304  
|||||  
258 GGGTATATTATTAACACTGGAAGAACAACTTTGATTATGGAACATACATGCTAAGTGGGGA 317  
|||||  
305 acaggtgtgtctaaatctgcaaaagcatctcccggaagagactgacaatgagataaagaact 364  
|||||  
318 ACAAGTGTGCAAAAATTTGCAAGACATTTGCTGGAACGATTAACGAGATTAATAAAACT 377  
|||||  
365 tctggagagactagatcacaaagcacatcaagaagcagaggtcttctctgcgtca 419  
|||||  
378 ACTGGAGGACTAGGATTGCAAGACCATTTAAGCAAGAGAAACATGAATGAGACA 432  
|||||

RESULT 2

AM928296 592 bp mRNA linear EST 18-MAY-2001  
LOCUS  
DEFINITION EST307029 tomato flower buds 8 mm to pre-anthesis, Cornell University Lycopersicon esculentum cDNA clone CTCIB20 5', mRNA sequence.

ACCESSION AM928296  
VERSION AM928296  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 592)  
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronling,C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.  
TITLE Generation of ESTs from tomato flower tissue  
JOURNAL Unpublished (1999)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.  
Location/Qualifiers  
1..592  
/organism="Lycopersicon esculentum"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CTCIB20"

FEATURES  
source

/clone\_lib="tomato flower buds 8 mm to pre-anthesis, Cornell University"  
/tissue\_type="flower"  
/tissue="buds 8mm-to-preanthesis"  
/dev\_stage="buds 8mm-to-preanthesis"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: Tanksley; flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

BASE COUNT 220 a 90 c 120 g 162 t

Query Match 28.5%; Score 245.4; DB 9; Length 592;  
Best Local Similarity 74.5%; Pred. No. 1.7e-44;  
Matches 309; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

ORIGIN

5 cgtctcgtttctctccaaagctgacatgagcaagaagccagagccagacagtgtaagt 64  
|||||  
18 CATCTTTTCTTTTAAAAAATATATAAAAAAGATGATGATTAATGCACT 77  
|||||  
65 cccaagatgtcgagtgagaaaagccgctgagacatgagaaagagatcatccatca 124  
|||||  
78 CTCAGATGTTGAAGTGAAGAAAGGCGCTTGACATGGAAGAAGATTATTTCTCATTA 137  
|||||  
125 actacatacgaaatcacagcggaagcagttggaactccctagccaaagctgctgtctaa 184  
|||||  
138 ACTCATTTGCTTAATCATGTGTAAGGTGTGTAACCTCTAGCTAAATCTGCTGCTCA 197  
|||||  
185 aaagtaacgggaagagtgctgcgtccgctgagactctctgcgacccgagctccgga 244  
|||||  
198 AACCTACTGGAAGAAAGTTGTAAGACTCCGATGGCTAAATTTCTTGACCTGATGTGCA 257  
|||||  
245 gaagcaacatcacactactgaggaagcaagctcctgatactgagacatgagcaagctggagaa 304  
|||||  
258 GGGTATATTATTAACACTGGAAGAACAACTTTGATTATGGAACATACATGCTAAGTGGGGA 317  
|||||  
305 acaggtgtgtctaaatctgcaaaagcatctcccggaagagactgacaatgagataaagaact 364  
|||||  
318 ACAAGTGTGCAAAAATTTGCAAGACATTTGCTGGAACGATTAACGAGATTAATAAAACT 377  
|||||  
365 tctggagagactagatcacaaagcacatcaagaagcagaggtcttctctgcgtca 419  
|||||  
378 ACTGGAGGACTAGGATTGCAAGACCATTTAAGCAAGAGAAACATGAATGAGACA 432  
|||||

RESULT 3

A1897681 519 bp mRNA linear EST 18-MAY-2001  
LOCUS  
DEFINITION EST267124 tomato ovary, TAMU Lycopersicon esculentum cDNA clone CLEJ30623, mRNA sequence.

ACCESSION A1897681  
VERSION A1897681  
KEYWORDS EST.  
SOURCE tomato.  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 519)  
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vasion,T., Holt,I.E., Liang,F., Upton,J., Ronling,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.  
TITLE Generation of ESTs from tomato carpel tissue  
JOURNAL Unpublished (1999)  
COMMENT Contact: CUGI  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
5 prime sequence.

FEATURES		Location/Qualifiers
source		1..519
		/organism="Lycopersicon esculentum"
		/cultivar="TA96"
		/db_xref="taxon:4081"
		/clone="cLED30G23"
		/clone_lib="tomato ovary, TAMU"
		/lissue_type="carpel"
		/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
		/lab_host="X11-Blue MRF"
		/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cLED - Tomato Carpel EST Library. OligoT-primed and directionally cloned cDNA in vector Lambda Zap II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT		193 a 73 c 113 g 140 t
ORIGIN		
Query Match	28.3%	Score 243.8; DB 9; Length 519;
Best Local Similarity	74.9%	Pred. No. 3.9e-44;
Matches 305; Conservative	0;	Mismatches 102; Indels 0; Gaps 0;
Oy	13	ttctctccaagctgacacatggacaagaagccagacgcagcagcagctgtaagtcaccagat 72
Db	27	TTTTTTTAAAAAATAATTAATAAAAAAAGATGATTAATAATGCAACCTCTCAACAT 86
Oy	73	gtcgcggtgagaaaaagggccgtagcagcttggaagaggtctcatctcattcaacata 132
Db	87	gttgaaatggagaaagggcccttgacactatggagaagaatttaattctcatTAACAT 146
Oy	133	gcgaatcacgycggaagcagcttgagaccctccatgaccaagctgctgtctaaacgtacc 192
Db	147	gctaatTCATGCGGAAGGtGTTTGAACCTCTAGCTAAATCTGCGTGTCTCAAGSTACT 206
Oy	193	gggaagagctgctgcgcctccgctgctgtaactctgagaccgcagctccgcgaaggaac 252
Db	207	GGAAAAAGCTTGTGAGACCTCCGATGCGTAAATATCTTCACACCGATGTCAGAGCGGCTAAT 266
Oy	253	atcactactgaagagcagctctcctatctatgtagacgacatgccaaagtgggaagaagctg 312
Db	267	ATTACACCTGGAAGCAACACTTTTGATTATGGACACTACATGCTAACTGGGGAACAAAGTGG 326
Oy	313	tctaaattgcacaagcatctctccgcgaagagactgacaatlgagataaagaactctggaag 372
Db	327	TCAAAAATTGGCAACCATTTGGCTGTGGAAGAACGGAATTAACGACATTAATAAAACCTAC 386
Oy	373	actagaatccaaaagcacatcaagcaagcagaagccttctctcgtca 419
Db	387	ACTAGATTTCAGAAAGCACATTTAAGCAAGAGCAACATGAATGAGACA 433
RESULT 4		
LOCUS	BE057370	
DEFINITION	BE057370	476 bp mRNA linear EST 03-DEC-2001
ACCESSION	BE057370	
VERSION	BE057370.1	
KEYWORDS		
SOURCE		
ORGANISM		soybean.
		glycine max
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE		1 (bases 1 to 476)
AUTHORS		Shoemaker,R., Kelm,P., Vodkin,L., Expelding,J., Coryell,V., Khanna,A., Bolla,B., Morris,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelshing,B., Allen,M., Bowers,R., Pearson,B., Swaller,T., Gibbons,M., Pepe,D., Harvey,N., Schurr,J., Rittler,E., Kohn,S., Shin,T., Jackson,V., Cardenas,M., McCann

TITLE  
JOURNAL.  
COMMENT

R., Waterston R. and Wilson, R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
444 Forest Park Parkway, Box B501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

This clone is available through: RegGen, Invitrogen Corp, 2130 South Memorial Parkway Huntsville, AL 35801 FOR further information call: (800)-933-4363 or contact via email: cshen@reggen.com  
Insert Length: 909 Std Error: 0.00  
Seq primer: -40RP from GIDCO  
High quality sequence stop: 420.

FEATURES  
Source location/qualifiers

1..476  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: gm-cl015-83r83"  
/clone.lib="gm-cl015"  
/tissue\_type="Mature flowers, field grown plants"  
/\_lab\_host="XLI10-Gold"  
/note="vector: pluescript II XR; Site\_1: EcoRI; Site\_2:  
xhoI; This cdna library was constructed from mRNA isolated  
from mature flowers of field grown plants. The CDNA  
library was prepared using the Stratagene pluscript II  
XR cdna library construction kit. Complementary DNA was  
synthesized from mrna using a primer consisting of a poly  
(dT) sequence with a xhoI restriction site. EcoRI adapters  
were ligated to the blunt-ended cdna fragments followed by  
xhoI digestion. The cdna fragments were directionally  
cloned into the EcoR-XhoI restriction site of the  
pluescript vector. The ligated cdna fragments were  
transformed into XLI10-Gold host cells. This library was  
constructed by Dr. Randy Shoemaker and Dr. John  
Erdelding."

BASE COUNT      145 a    113 c    116 g    102 t  
ORIGIN

Query Match                  28.1%; Score 242; DB 9; Length 476;  
Best Local Similarity         74.6%; Pred. No. 9.9e+44;

Matches 337; Conservative    0; Mismatches 100; Indels 15; Gaps 2,

Gy    61 aagtcccaaatgtctcgaggtagaaaaggccgtygaagaatgggatcatcttc    120  
||| |||||||  
Db    16 AGCTTCATGAACCGAAGTAGAAAAGGGCCCATGTGCACAATGAAGAACHTAATTATTTG    75  
  
Gy    121 atcaactacatacgcgaattaccgagaaagcgattgsaacctcttaagccaagcttgct    180  
|||| |  
Db    76 ATCACTTAAATTATTTGCCAAATCAGGGAGGGGTTTTGGAACCTCTTTGGCCAAGACTGCTTGA    135  
  
Gy    181 ctanaaacgtaccggygaagatgtctcgcccgcgtygaccttatctggaaccgacytic    240  
||| |||||||  
Db    136 CTTAAACGTAACCAGGAAGAGTGTCGGCGCTCCGGTGCTTAACCTACCTCGCTGATGTT    195  
  
Gy    241 cgagagagcacatatctaactctgaagagcaagctcctgatcataatgaatcyttcaaagytg    300  
|||| |  
Db    196 ACAAGAGGGGAATAATTAACACCCAGAGAACAGCTTTTGGATCATGTGAACCTTATCAATCAAGTNG    255  
  
Gy    301 ggaaaaagatgctgtaaattgacaagcaatcttcocgsgaagaagacagajataagaataag    360  
|||| |  
Db    256 GGAAACAGGTGCTCAAATTTGCCAAGCAAGCTTAACCCGGAAGAGCTGATTAATAGATTAAAG    315  
  
Gy    361 aaactcttgagagactagaatccaanaagcacatcaagcaagcagaagcttctc-----    412  
|||| |  
Db    316 AACTACTGGAGAGCAAGAAGTCCAGAAAGCACGCTCAAGCAAGCTTCCAGCAGCTTCCAGCA    375  
  
Gy    413 -ctgtcagagctccgagatgaagtgt-----caagcaagaacaaagacatatgtgccagc    465  
|||| |  
Db    376 CAGAGTAGTAATTCGAGTAGTAATTTATCATCCCCAAGCTTGGACATAGGCCAAGCTGCCAC    435

```

QY      466 atgcagagccgcatgagacactacacacc 497
      ||| | ||| | ||| | ||| | ||| | ||| |
Db      436 ATGCGCGACGCCCATGAAACCTATTCTCCACC 467

RESULT  5
LOCUS   BE805071
DEFINITION BE805071 395 bp mRNA linear EST 06-DEC-2001
          ss37407.y1 Gm-c1061 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
          Gm-c1061-877 5' similar to TR:Q39028 Q39028 ATMYB2.1, mRNA
          sequence.
ACCESSION BE805071
VERSION   BE805071.1 GI:10236183
KEYWORDS  EST.
SOURCE    soybean.
          Glycine max
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
REFERENCE 1 (bases 1 to 395)
AUTHORS   Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corvett,V., Khanna
          ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
          Wylie,T., Underwood,K., Stepcie,M., Theising,B., Allen,M., Bowers
          ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
          ,R., Rittner,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
          ,R., Waterston,R. and Wilson,R.
          Public Soybean EST Project
          Unpublished (1999)
TITLE     Shoemaker R./Public Soybean EST Project
JOURNAL   Public Soybean EST Project
COMMENT   Contact: Shoemaker R./Public Soybean EST Project
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: estewatson.wustl.edu
          This clone is available through: Resgen, Invitrogen Corp. 2130
          South Memorial Parkway Huntville, AL 35801 For further information
          call: (800)-533-4363 or contact via email: ccuteresgen.com
          High quality sequence stop: 367.
FEATURES
source    1..395
           /organism="Glycine max"
           /db_xref="taxon:3847"
           /clone="GENOME SYSTEMS CLONE ID: Gm-c1061-877"
           /clone_1lb="Gm-c1061"
           /tissue_type="mature flowers of field grown plants"
           /lab_host="DH10B"
           /note="Vector: pBluescript II SK-; Site_1: EcoRI, Site_2:
           XhoI; The cDNA library was constructed from mRNA isolated
           from mature flowers of field grown plants for the cultivar
           Raiden. Complementary DNA was synthesized from mRNA using
           a primer consisting of a poly(dT) sequence with a XhoI
           restriction site. EcoRI adapters were ligated to the
           blunt-ended cDNA fragments followed by XhoI digestion. The
           cDNA fragments were directionally cloned into the
           EcoRI-XhoI restriction site of the pBluescript vector. The
           ligated cDNA fragments were transformed into DH10B host
           cells (GibcoBRL). This library was constructed in the
           Laboratory of Dr. Randy Shoemaker."

BASE COUNT 127 a 86 c 103 g 79 t
ORIGIN

```

```

Query Match      28.1% Score 241.8; DB 10; Length 395;
Best Local Similarity 81.8%; Pred. No. 1.1e-43;
Matches 279; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

QY      61 aagtcaccaagatgctgagaggaagagccgtgagcagatggaagagatctcatctc 120
      ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      33  ACgtcTCATGATCTGAAGTGAAGAGGGCGCATGACATGCAAGAACTTAATCTTG 92

```

```

QY      121 atcaactacatagcgaatcacagcggaaggaattggaactccctagcacaagctgctg 180
      ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      93  ATCACCCTATTTCCATATACCGGGAGGGGTTTGGAACTCTTGGCACAAGGCTGCTGGA 152

QY      181 ctataacgctacgggaagagctgctcgctcgctgctgaactatctcgaccgcagctc 240
      ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      153  CTTAAACGTACCGGAAGATTTGCCGCTCCGCTCAACTACCTCCCTCTGTTGTT 212

QY      241 cggagagcgacaactcaactagagcagcctcgtatcatgtgaactgctccaagt 300
      ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      213  AGAAGAGCGAATTTTACACCGCAGACACGCTTTTATATATGAACTTATGCAAAAGTGG 272

QY      301 ggaacaggtgctctaaatctgcaagcatctcccggaagagctgacaaatgaglaaag 360
      ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      273  GGAACAGGCTGCTCAAAATTTGCCAAGCATCTACCCGGAAGCACTGATATGATTAAG 332

QY      361 aactcttgagagctagaatccaaagacatataagcaagc 401
      ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      333  AACTACTGAGAGACAGATCCAGACACTTCACCAAGC 373

```

```

RESULT  6
LOCUS   AI930997
DEFINITION AI930997 523 bp mRNA linear EST 30-NOV-2001
          sb43h07.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
          Gm-c1015-278 5' similar to TR:Q39028 Q39028 ATMYB2.1, mRNA
          sequence.
ACCESSION AI930997
VERSION   AI930997
KEYWORDS  EST.
SOURCE    soybean.
          Glycine max
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
          Glycine.
REFERENCE 1 (bases 1 to 523)
AUTHORS   Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corvett,V., Khanna
          ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
          Wylie,T., Underwood,K., Stepcie,M., Theising,B., Allen,M., Bowers
          ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
          ,R., Rittner,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
          ,R., Waterston,R. and Wilson,R.
          Public Soybean EST Project
          Unpublished (1999)
TITLE     Shoemaker R./Public Soybean EST Project
JOURNAL   Public Soybean EST Project
COMMENT   Contact: Shoemaker R./Public Soybean EST Project
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: estewatson.wustl.edu
          This clone is available through: Resgen, Invitrogen Corp. 2130
          South Memorial Parkway Huntville, AL 35801 For further information
          call: (800)-533-4363 or contact via email: ccuteresgen.com
          Seq primer: -40RP from Gibco
          High quality sequence stop: 422.
FEATURES
source    1..523
           /organism="Glycine max"
           /db_xref="taxon:3847"
           /clone="GENOME SYSTEMS CLONE ID: Gm-c1015-278"
           /clone_1lb="Gm-c1015"
           /tissue_type="mature flowers, field grown plants"
           /lab_host="XL10-Gold"
           /note="Vector: pBluescript II XRI; Site_1: EcoRI, Site_2:
           XhoI; This cDNA library was constructed from mRNA isolated
           from mature flowers of field grown plants. The cDNA
           library was prepared using the Stratagene pBluescript II
           XR cDNA library construction kit. Complementary DNA was
           synthesized from mRNA using a primer consisting of a poly
           (dT) sequence with a XhoI restriction site. EcoRI adapters
           were ligated to the blunt-ended cDNA fragments followed by

```





KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

REFERENCE Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W., and May, G.D.  
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Trifoliaceae: Medicago.

1 (bases 1 to 688)

AUTHORS

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula flower library

JOURNAL

COMMENT Unpublished (2001)  
Contact: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Insert Length: 688 Std Error: 0.00  
Plate: 091 row: A column: 12  
Seq primer: TCACACAGGAACACCTATGCAC.

FEATURES

source

1..688  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF091A12FL"  
/clone\_1lb="Developing flower"  
/tissue\_type="Developing flowers"  
/dev\_stage="Developmentally pooled. Contains a mixture of very young, developing, fully-opened flowers and flowers in early transition into pods."  
/note="Vector: Lambda Zap; cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the GigaPack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using Exassist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

BASE COUNT 235 a 144 c 125 g 182 t 2 others  
ORIGIN

Query Match 27.6%; Score 238.2; DB 10; Length 688;  
Best Local Similarity 79.2%; Pred. No. 6,9e-43;  
Matches 282; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

50 acgacatgtgtaagtcctccagatctcgagctgagaaagcgctgagcatggaagag 109  
111  
59 AAGAGTCGAGTCTTTCACAAAGATCTGATATACGAAAGGACCATGACATGGAAAG 118  
110 atctcatcctcaatacatatagcgaaatcaagcgagctggaactccctagcca 169  
119 ACTTGATCTTGATCACTATATTTGCCAATCATGNGNAAGCTGTTTGAACCTCTTGCTGA 178  
170 aagctgctgctctaaacgtaacgtaacgtaacgctgctgctcgctgagcgtggaataatctgc 229  
179 AATCTGCTGCTCTTAAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAAC 238  
230 gaccgcagctcgcgagagggcacatcaactactactagagagcagctctcgatcatatggaactgc 289  
239 GACCGGATGTTAAGACGAGGGAATATTACACCTGAGCAACACTCTTGATATGAACTTTC 298  
240 atgcacagctgagaaacagctgctaaatctgaaagcactctcccgagagagagcagaca 349  
299 ATGCTAACTGGGAAACGGGTGCGAANAATTCGGAACATCTTACCGGGAAGAACGATGA 358  
350 atgagataaagaaactctcgagagactagaaatccaaagacatcaagcaagcagag 405

Db 359 ATGACATTAAGACTACTGAGAGACAGCATCCAAAGCATATCAAGCAAGCTGAG 414

RESULT 9

BE324639

LOCUS 558 bp mRNA linear EST 21-DEC-2000

DEFINITION NF024C04P1.F1023 Phosphate starved leaf Medicago truncatula cDNA

clone NF024C04P1.5, mRNA sequence.

ACCESSION BE324639.2

VERSION GI:11965871

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..558  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF024C04PL"  
/clone\_1lb="Phosphate starved leaf"  
/tissue\_type="leaf"  
/dev\_stage="trifoliolate"  
/note="Vector: Lambda Zap; At the trifoliolate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this 30 day period, the plants were fertilized twice weekly with 1/2 Hoaglands solution containing only 20um potassium phosphate. RNA was prepared from above ground tissues."

BASE COUNT 196 a 112 c 110 g 140 t  
ORIGIN

Query Match 27.6%; Score 237.6; DB 10; Length 558;  
Best Local Similarity 79.2%; Pred. No. 9,4e-43;  
Matches 282; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

50 acgacatgtgtaagtcctccagatctcgagctgagaaagcgctgagcatggaagag 109  
111  
60 AAGAGTCGAGTCTTTCACAAAGATCTGATATACGAAAGGACCATGACATGGAAAG 119  
110 atctcatcctcaatacatatagcgaaatcaagcgagctggaactccctagcca 169  
120 ACTTGATCTTGATCACTATATTTGCCAATCATGNGNAAGCTGTTTGAACCTCTTGCTGA 179  
170 aagctgctgctctaaacgtaacgtaacgtaacgctgctgctcgctgagcgtggaataatctgc 229  
180 AATCTGCTGCTCTTAAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAAC 239  
230 gaccgcagctcgcgagagggcacatcaactactactagagagcagctctcgatcatatggaactgc 289  
240 GACCGGATGTTAAGACGAGGGAATATTACACCTGAGCAACACTCTTGATATGAACTTTC 299

OY 290 atgccagtcggggaacagtggtctcaaaatgcgaagcatctcccggaagactgaca 349  
|||||  
Db 300 ATGCTTAAGTGGGGAACAGACTGCTCGAATATTGCCAATATCTACCGGGAAGACCGATA 359  
OY 350 atgagaataagaactctctggagactagataccaagaacatcaagcagaag 405  
|||||  
Db 360 ATGAGATTAAGAAGACTCTCTGAGAGACAGATCCAAAAGCATATCAAGCAAGCTGAG 415

RESULT 10  
BG457971 669 bp mRNA linear EST 19-MAR-2001  
LOCUS NF037A10P1.F1070 Phosphate starved leaf Medicago truncatula cDNA  
DEFINITION clone NF037A10P1 5', mRNA sequence.  
ACCESSION BG457971  
VERSION BG457971.1 GI:13381296  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
REFERENCE 1 (bases 1 to 669)  
AUTHORS Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores  
, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.  
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
JOURNAL Medicago truncatula phosphate-starved leaf library  
COMMENT Unpublished (2000)  
Contact: Harrison MJ  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7325  
Fax: 580 221 7380  
Email: mjharrison@noble.org  
Insert Length: 669 Std Error: 0.00  
Plate: 037 row: A column: 10  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. 669  
FEATURES  
source  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF037A10P1"  
/clone\_lib="Phosphate starved leaf"  
/tissue\_type="leaf"  
/dev\_stage="trifoliolate"  
/note="Vector: Lambda Zap: At the trifoliolate stage, M.  
truncatula plants were transplanted to phosphate-free sand  
and grown for a further 30 days. During this 30 day  
period, the plants were fertilized twice weekly with 1/2  
Hoaglands solution containing only 20uM potassium  
phosphate. RNA was prepared from above ground tissues."

BASE COUNT 225 a 141 c 127 g 176 t  
ORIGIN

Query Match 27.6%; Score 237.6; DB 10; Length 669;  
Best Local Similarity 79.2%; Pred. No. 9.3e-43;  
Matches 282; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 50 acgacagtggttaagtcaccaagatgctgaggtgagaaaagggcggtgacgataggaag 109  
|||||  
Db 60 AAGAGTCACTTCTTCACAAATCCTGATGTACGAAAGGACCATGTGACATGGAAG 119  
OY 110 atctcacccatcaatcatagcgaatcagcgaagcgatggaatccctagcca 169  
|||||  
Db 120 ACTTGATCTTGATCACTAATATTTGCCATATGATGTAAGGTGTTGGAACTCTTGCTTA 179  
OY 170 aagctgcgtgtctaaacagctaccggaagagtgctgcgcgcggtgctgaactatctgc 229  
|||||  
Db 180 AATCTGCTGGCTTTAAACGTACCGAAAGAGTTGACAGGCTTCGCTGCTAATATCTTTC 239

OY 230 gaccgcagctccggagaggaacacatcactactaggagagcaagctctgatactgaaactgc 289  
|||||  
Db 240 GACCGGAGTGTAGACAGAGGAATATTACCTGAGGAACAACCTCTTGATATTTGAATTC 299  
OY 290 atgccagtcggggaacagtggtctcaaaatgcgaagcatctcccggaagactgaca 349  
|||||  
Db 300 ATGCTTAAGTGGGGAACAGACTGCTCGAATATTGCCAATATCTACCGGGAAGACCGATA 359  
OY 350 atgagaataagaactctctggagactagataccaagaacatcaagcagaag 405  
|||||  
Db 360 ATGAGATTAAGAAGACTCTCTGAGAGACAGATCCAAAAGCATATCAAGCAAGCTGAG 415

RESULT 11  
B1273011 447 bp mRNA linear EST 18-JUN-2001  
LOCUS B1273011  
DEFINITION NF097F02FL1026 Developing flower Medicago truncatula cDNA clone  
NF097F02FL 5', mRNA sequence.  
ACCESSION B1273011  
VERSION B1273011.1 GI:14882835  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
REFERENCE 1 (bases 1 to 447)  
AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.  
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
JOURNAL Medicago truncatula flower library  
COMMENT Unpublished (2001)  
Contact: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Insert Length: 447 Std Error: 0.00  
Plate: 097 row: F column: 02  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. 447  
FEATURES  
source  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF097F02FL"  
/clone\_lib="Developing flower"  
/tissue\_type="Developing flowers"  
/dev\_stage="Developmentally pooled. Contains a mixture of  
very young, developing, fully-opened flowers and flowers  
in early transition into pods."  
/note="Vector: Lambda zap: cDNA was prepared from polyA+  
enriched, pooled samples of equivalent amounts of total  
RNA from very young, developing, fully-opened flowers and  
flowers transitioning into pods. The cDNA was  
directionally ligated into the Uni-Zap XR vector  
(Stratagene) and packaged using the GigaPack III Gold  
packaging extracts. Phagemids containing cDNA inserts were  
in vivo excised from the recombinant Uni-Zap XR vector  
using ExAssist helper phage and the E. coli strain  
XL1-Blue MRF' (Stratagene). Excised plasmids were plated  
using SOLR cells."

BASE COUNT 158 a 85 c 85 g 119 t  
ORIGIN

Query Match 27.4%; Score 236.4; DB 10; Length 447;  
Best Local Similarity 74.6%; Pred. No. 1.7e-42;  
Matches 297; Conservative 0; Mismatches 101; Indels 0; Gaps 0;





## REFERENCE

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
Lycopersicon.  
1 (bases 1 to 371)

## AUTHORS

Alcala,J., Vredalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,  
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,  
C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,  
S.D. and Giovannoni,J.  
Generation of ESTs from tomato carpel tissue  
Unpublished (1999)

## TITLE

On Mar 8, 1999 this sequence version replaced gi:4383294.

## JOURNAL

COMMENT

contact: CUCI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

FOLVA=NO.

## FEATURES

Location/Qualifiers

1..371

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="CLEDI9F4"

/clone\_lib="tomato ovary, TAMU"

/tissue\_type="carpel"

/dev\_stage="5 days pre-anthesis to 5 days post-anthesis"

/lab\_host="XLI-Blue MR"

/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:

XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and

directionally cloned cDNA in vector Lambda Zap II with 5'

and 3' ends located at the EcoRI and XhoI sites,

respectively."

BASE COUNT 129 a 53 c 78 g 111 t  
ORIGIN

## Query Match

23.6% Score 203.2; DB 9; Length 371;

Best Local Similarity 74.4%; Pred. No. 4,2e-35;

Matches 256; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
QY 13 ttctctccaaagcgcacatcgacaagacgcagacagctggtaagtcaccaagat 72
DB 28 TTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 87
QY 73 gtcgaggtgagaaaggccgctgagacgacgacgacgacgacgacgacgacgac 132
DB 88 GTTGAAGTGAGGAAGGCGCTGAGCTATGGAAGAGATTAAATTCATTAACTACAT 147
QY 133 ggcgatacagcgagacgacgacgacgacgacgacgacgacgacgacgacgac 192
DB 148 GCTAATCATGTGTAAGGTGTTGGAACCTCTGACTTAATCTGCTGCTCAACGTACT 207
QY 193 ggcgagagcttcgagctcgctgacgacgacgacgacgacgacgacgacgac 252
DB 208 GGAATAAGCTGTACACTCCGATGCTAATTAATTAATTAATTAATTAATTAAT 267
QY 253 atcactcctgagagacgacgacgacgacgacgacgacgacgacgacgacgac 312
DB 268 ATTACACTGAGAAACAATTGATTATGAACTACATGCTAAGTGGGGAACAAGTGG 327
QY 313 tctaaatctgaaagcactctcccgaaagacgacgacgacgacgacgacgac 356
DB 328 TCAGAAATTCGAAGCATTTGCTGGAAGAGAGATTAACGAGAT 371
```

Search completed: August 15, 2002, 03:33:55  
Job time: 6260 sec

R; Theologis, A.; Ecker, J. R.; Palm, C. J.; Federspiel, N. A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C:Date: 11-Jun-1999 #sequence\_revision 19-Jun-1999 #text\_change 22-Oct-1999  
 R:Authors: Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A66141; MIMD:21016719  
 A:Accession: G86145  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1284 <STO>  
 A:Cross-references: GB:AE005172; NID:g8920588; PIDN:AF81310.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 38.6%; Score 576; DB 2; Length 1284;  
 Best Local Similarity 50.0%; Pred. No. 1,2e-35;  
 Matches 129; Conservative 31; Mismatches 54; Indels 44; Gaps 8;

OY 26 MALPG---LAPNLSSASVSASAS-----EDSAKTRKRYTTTKRSFESQEDHKEFLA 76  
 DB 1013 MSLEPGNTLPHTATTPVSIKRNKTSFEDPTKVKRKYTTTKRSKEMTFQEDHKEFLA 1072  
 OY 77 LQLEPDMKKIEAFVGSKVIVQIRSHAKYFLKVKNGTREVHPKPKKASHPYPOKA 136  
 DB 1073 LHLFEDDMKKIKAFVGSKVIVQIRSHAKYFLKVKNGTREVHPKPKKASHPYPOKA 1132  
 OY 137 SKNVPVSOQVSTAFPTAAT--QLDSGYPRAESSILTKSSSCPTVSSVNHHTIPSIDAF 196  
 DB 1133 PK-----FTLSSNALFQHDYLY-----WNSSHNVISTTRHGLVHCDVSI 1173  
 OY 197 ---VEKDDGCPGCIETGNKSSGTSSTSPPT-----WPCSEIPEKVKPDSQVYKF 245  
 DB 1174 PSSVYIKERK---VSENCSTSSSRDKQRTRIYETINDQESGCK-PHNAVPRFAEVNPF 1228  
 OY 246 IGSVPDPSTTDHLKKE 263  
 DB 1229 IGSVPDPKTTGHWKRLKE 1246

RESULT 3  
 T48305  
 Hypothetical protein F9G14.150 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T48305  
 R:Revan, M.; Terry, N.; Ardiles, W.; Buysaert, C.; Dasseville, R.; De Clerck, R.; De  
 ewen, H.W.; Budd, S.; Lemcke, K.; Meyer, K.F.X.  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z24491  
 A:Accession: T48305  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-307 <BEV>  
 A:Cross-references: EMBL:AL162973  
 A:Experimental source: cultivar Columbia; BAC clone F9G14  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 65/2; 84/3; 126/1; 193/1; 235/1; 271/3; 291/3  
 A:Note: F9G14.150

Query Match 38.0%; Score 567.5; DB 2; Length 307;  
 Best Local Similarity 44.8%; Pred. No. 8.6e-36;  
 Matches 133; Conservative 17; Mismatches 62; Indels 85; Gaps 6;  
 OY 5 AETRYDSFETTSQSSVDVGMALPGIALPNLSSASVSASSEDSAKTRKRYTTTKRSRES 64  
 DB 14 AETSTDAETFT-----IATTEAGAEAKKVRKAYTTTKRSRES 50

OY 65 MSEOEDKLEALQLEPDMKKIEAFVGSKVIVQIRSHAKYFLKVKNGTREVHPKPK 124  
 DB 51 WTEGEHKEFLALQLEPDMKKIEDPFGSKVIVQIRSHAKYFLKVKNGTREVHPKPK 110  
 OY 125 KRKASHPYPOKASKNVPSQVSTAFPTAAT--QLDSGYPRAESSILTKSSSCPTVSS 184  
 DB 111 KRKAHPYPQKASKNVPSQVSTAFPTAAT--QLDSGYPRAESSILTKSSSCPTVSS 164  
 OY 185 VNHHTIPSIDAFVEKDDGCPGCIETGNKSS-----GSTESSPTWP 226  
 DB 165 I-----PPEDLDITCGABGRSHLYSVIFCFVDSGNDMISERSP 206  
 OY 227 PCSEIPEKVK-----PDSQVYKFIQSVDPSTTDHLKKE 263  
 DB 207 SASGIGSSSRLLSDSKGLRLAKQAFSMHGLPDPFAVEVNFISVDPDPSKGRMKLKE 263

RESULT 4  
 T01715  
 Hypothetical protein A\_I6002N01.20 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 22-Oct-1999  
 C:Accession: T01715  
 R:Scheel, P.; Magel, L.  
 submitted to the EMBL Data Library, June 1997  
 A:Description: The sequence of A. thaliana I6002N01.  
 A:Reference number: Z14407  
 A:Accession: T01715  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-213 <SCH>  
 A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191140; ATSP:A\_I6002N01.20  
 C:Genetics:  
 A:Gene: ATSP:A\_I6002N01.20  
 A:Introns: 76/2; 95/3; 137/1; 192/1

Query Match 33.7%; Score 502.5; DB 2; Length 213;  
 Best Local Similarity 56.4%; Pred. No. 4.8e-31;  
 Matches 106; Conservative 20; Mismatches 45; Indels 17; Gaps 5;

OY 22 DLVGNALP-----GLAPNLSSASVSASSEDSAKTRKRYTTTKRSFESQEDHKEFLA 76  
 DB 15 DSSNMSLPSSDGFSGIPATGRFS-TVSESEDPYTKIRKRYTTTKRSKEMTFQEDHKEFLA 73  
 OY 77 LQLEPDMKKIEAFVGSKVIVQIRSHAKYFLKVKNGTREVHPKPKKASHPYPOKA 136  
 DB 74 LHLFEDDMKKIKAFVGSKVIVQIRSHAKYFLKVKNGTREVHPKPKKASHPYPIKA 133  
 OY 137 SKNVPVSOQVSTAFPTAAT--QLDSGYPRAESSILTKSSSCPTVSSVNHHTI----P 190  
 DB 134 PKNV-----AYTSLPSSSTPLPLEPGYIYSDSKSLMGNAVCASTSSWNHSTNLKPK 188  
 OY 191 STDASFVE 198  
 DB 189 VIEGSFTD 196

RESULT 5  
 T02684  
 MYB-related transcription factor (CCA1) [imported] - *Arabidopsis thaliana*  
 N:Alternate names: DNA-binding protein (CCA1), protein F19D11.11  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
 C:Accession: T02684; G84907  
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;  
 submitted to the EMBL Data Library, September 1998  
 A:Description: *Arabidopsis thaliana* chromosome II BAC F19D11 genomic sequence.  
 A:Reference number: Z14698  
 A:Accession: T02684  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA





Best Local Similarity 27.3%; Pred. No. 0.0028;  
Matches 69; Conservative 30; Mismatches 100; Indels 54; Gaps 13;

```

OY 12 FETTSGGSSVDLVGMA--LPGLAPN--LSSASVSASASED-----SAKKIRPYITTKS 61
      | : | | | | : | | : | : | | | | | : | |
DB 47 FASGSGGSGTPLVNGVYHDSGDPDPDHAVGSGADCVASDVFVAGSSSSNER-----KK 100
      | : | | | | : | | : | : | | | | | : | |
OY 62 RESSEGEHKLFEAL-QLFDRDMKTI-EAFVSGKTYIQRSHAKYFLKYNQKREHY 119
      | : | | | | : | | : | : | | | | | : | |
DB 101 GVPTEEEHMFLLGLCKLCKGDMRGIAARNYVISRPTQVASHAKYFIR-QSNMSR--- 156
      | : | | | | : | | : | : | | | | | : | |
OY 120 PPPPRKASHPYPOKASKNVP-VSQOVSTAPPAATQDLSGYR-----R 164
      | : | | | | : | | : | : | | | | | : | |
DB 157 ---RKRSSSLFDIYADSGGTPPMVSRDFLADDPAAQAEKMSNLLPPTPAVDECESSNGSA 213
      | : | | | | : | | : | : | | | | | : | |
OY 165 AESSSILTKSGSSCPYVSSVWHHTIPSIDASFYEKDDGCPGLETGNCCSGSTESSPPT 224
      | : | | | | : | | : | : | | | | | : | |
DB 214 ASANSIDGHALPIPESSQYQH---PLVYPAYV-----APFPMYPMPYGTAE----- 260
      | : | | | | : | | : | : | | | | | : | |
OY 225 WPPCSEPEKVKP 237
      | : | | | | : | | : | : | | | | | : | |
DB 261 -PAIAETHEVLKP 272
      | : | | | | : | | : | : | | | | | : | |

```

## RESULT 9

I-box binding factor-like protein - Arabidopsis thaliana  
N:Alternate names: protein T1E3.120  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T48472  
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De  
ewes, H.W.; Rudd, S.; Lemcke, K.; Meyer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: 224491  
A:Accession: T48472  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-215 <BEV>  
A:Cross-references: EMBL:AL162972  
A:Experimental source: cultivar Columbia; BAC clone T1E3  
C:Genetics:  
A:Map position: 5  
A:Introns: 107/2  
A:Note: T1E3.120

Query Match 9.1%; Score 136.5; DB 2; Length 215;  
Best Local Similarity 29.3%; Pred. No. 0.0034;  
Matches 51; Conservative 24; Mismatches 60; Indels 39; Gaps 9;

```

OY 10 DSFEFTSGSSVDLVGMALPGLAPNLSASVSASASEDSAKKI-----RKPYITTKSRESW 65
      | : | | | | : | | : | : | | | | | : | |
DB 53 DVEFIDSG--RDVY-----PDVYMDSSAAAGWDSDAGQISFGSKHGESRRKRGTPW 101
      | : | | | | : | | : | : | | | | | : | |
OY 66 SEGEHKLFEALQLFDR-DWKKTIEA-FVSGKTYIQRSHAKYFLK---YQKNGTEHY- 119
      | : | | | | : | | : | : | | | | | : | |
DB 102 TENEHKLLFLGLKRYGCKDRSSISRNVYVTRPTQVASHAKYFLKQNSYKRRKRSSIH 161
      | : | | | | : | | : | : | | | | | : | |
OY 120 -----PPPRKASHPYPOKASKNVPVSOOVSTAPPAATQDLSGYR 162
      | : | | | | : | | : | : | | | | | : | |
DB 162 DITTVDATLAMPGSNMMDWTQHGSPVQA-----PQQQDINSEF---GQQLNPGHF 208
      | : | | | | : | | : | : | | | | | : | |

```

## RESULT 10

hypothetical protein F7J8.180 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 19-May-2000  
C:Accession: T45960  
R:Bevan, M.; Zimmermann, W.; Grunselmen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Lem  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: 223018

A:Accession: T45960  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-267 <BEV>  
A:Cross-references: EMBL:AL137189  
A:Experimental source: cultivar Columbia; BAC clone F7J8  
C:Genetics:  
A:Map position: 5  
A:Introns: 154/2  
A:Note: F7J8.180  
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat hom

Query Match 9.0%; Score 134; DB 2; Length 267;  
Best Local Similarity 24.9%; Pred. No. 0.0069;  
Matches 49; Conservative 30; Mismatches 76; Indels 42; Gaps 7;

```

OY 19 SSVDLVGMALPGLAPLSSASVS-----ASASDSAKKIRKPYT 57
      | : | | | | : | | : | : | | | | | : | |
DB 81 SDIEAGLIPPIPGIGDASANSDFYFGLNSSYGYDYVVGKRSPPATDCFRSPMPPEK 140
      | : | | | | : | | : | : | | | | | : | |
OY 58 ITKRESSEGEHKLFEALQLFDR-DWKKI-EAFVSGKTYIQRSHAKYFLKYNQKNGT 115
      | : | | | | : | | : | : | | | | | : | |
DB 141 ERKKGVPTWEDENLRFLMLCKYKGDWRNIAKSFYTTTPTQVASHAKYFLRQLTDC- 199
      | : | | | | : | | : | : | | | | | : | |
OY 116 REHVPPPRKRNAS-HPYQASKNVPVSOQVSTA-----PPTAATQDLSGYRPAES 167
      | : | | | | : | | : | : | | | | | : | |
DB 200 -----KDKRRSSIHDI---TTVINPDADASATATTADVALSPIPANSFDFLQPNPHY 249
      | : | | | | : | | : | : | | | | | : | |
OY 168 SSILTKSGSSCPYVSSW 184
      | : | | | | : | | : | : | | | | | : | |
DB 250 SFNASASASYNAFPQW 266
      | : | | | | : | | : | : | | | | | : | |

```

## RESULT 11

hypothetical protein F9E10.31 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H96777  
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hulmar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucero, J.S.; Matli, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A66141; MUID:21016719  
A:Accession: H96777  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-265 <STO>  
A:Cross-references: GB:AE005173; NID:96646780; PIDN:AAZ21092.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F9E10.31  
A:Map position: 1

Query Match 8.9%; Score 133.5; DB 2; Length 265;  
Best Local Similarity 25.4%; Pred. No. 0.0074;  
Matches 59; Conservative 26; Mismatches 66; Indels 81; Gaps 10;

```

OY 9 RDSFEFTSGSSVDLVGMALPGLAPNLSASVSASASEDSAKKIRKPYITTKSRSWSEQ 68
      | : | | | | : | | : | : | | | | | : | |
DB 65 QDMNKTPPAGYASND--ELP-----MSS-----SNCKIER-----KRGVPTEE 101
      | : | | | | : | | : | : | | | | | : | |
OY 69 EHDKFLFEALQLFDR-DWKKTIEA-FVSGKTYIQRSHAKYFLK----- 109
      | : | | | | : | | : | : | | | | | : | |
DB 102 EHKLLFLGLKRYGCKDMKISRNFVKTRTSQVASHAKYFLRNRNLRNRRRSSLPDMT 161
      | : | | | | : | | : | : | | | | | : | |

```



Db 195 IHDI---TTVNIPOSDPDAAAA 212

## RESULT 15

B45344

probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)

C:Species: suid herpesvirus 1

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999

C:Accession: B45344

R:Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzler, M.

Virology 179, 365-377, 1990

A:Title: Pseudotables virus immediate-early gene overlaps with an oppositely oriented OR

A:Reference number: A45344; MUID:91021039

A:Accession: B45344

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-1733 &lt;VLC&gt;

A:cross-references: GR:M4651; NID:9334070; PIDN:AAA47471.1; PID:9334072

C:superfamily: pseudotables virus 1 nuclear antigen

## Query Match

7.8%; Score 116.5; DB 1; Length 1733;

Best Local Similarity 26.8%; Pred. NO. 1.6;

Matches 38; Conservative 15; Mismatches 56; Indels 33; Gaps 6;

QY 116 REHVPPRPKRKASH--PYOKASKNV--PVSQOVSTAFPTATQUDSGY-----PRA 165  
 Db 93 KDHLPFRTRRDHQRPPPTTTTIKDPQHQPDPILLPTKTLQEDPHLLRPRDPPSA 152  
 QY 166 ESSSILTKSGSSCPTVSSWVHRTIPSIDASFVEKDDGPGIETGNCSGSGTESSPPT- 224  
 Db 153 KTHHHODPPGCGPPSTSSHHHODP-----PCGGPPSPPPRPSTSSSSHQGPST 203  
 QY 225 -----WPPCSEIPEKV 235  
 Db 204 RPPPPQRPPRPWPPPS--PQXI 223

Search completed: August 12, 2002, 10:11:07  
 Job time: 245 sec